

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN EDEBİYAT FAKÜLTESİ
MATEMATİK MÜHENDİSLİĞİ PROGRAMI



**İTÜ MATEMATİK MÜHENDİSLİĞİ ÖĞRENCİLERİNİN MEZUNİYET
BAŞARILARINI YÜKSELTMEYİ HEDEFLEYEN BİR VERİ
MADENCİLİĞİ UYGULAMASI**

BİTİRME ÖDEVİ

Nur Sena TANRIVERDİ 090080038

Teslim Tarihi: 31.05.2013

Tez Danışmanı: Yar. Doç. Dr. Ahmet KIRIŞ

MAYIS 2013

ÖNSÖZ

Bu çalışmayı hazırlamamda bana yol gösteren, yardımını, bilgisini ve her türlü desteği hiçbir zaman esirgemeyen Sayın Hocam Yar. Doç. Dr. Ahmet KIRIŞ'a, çalışmam için gerekli öğrenci verilerini elde etmemi sağlayan İTÜ Öğrenci Dekanlığı Dekanı Sayın Prof. Dr. Yılmaz TAPTIK'a ve Dekan Yardımcısı Sayın Hocam Yar. Doç. Dr. Semra AHMETOLAN'a ve İTÜ Öğrenci İşleri Daire Başkanlığı çalışanı Sayın Uzman Nehir SERTKAYA'ya ve öğrenim hayatım boyunca omuz omuza yürüdüğüm kıymetli arkadaşlarıma, bana sevgi, güven ve her türlü desteği veren biricik aileme en içten teşekkürlerimi sunarım.

Mayıs, 2013

Nur Sena TANRIVERDİ

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	1
1. Giriş	2
1.1. Açıklamalar	2
1.2. Öğretim Süresi ve Öğretim Planları	4
1.3. Notlar ve İşaretler	6
1.4. Ağırlıklı Genel Not Ortalaması ve Gözetim Durumu	7
2. VERİ MADENCİLİĞİ	9
2.1. Tanım	9
2.2. Kullanım Alanları	10
2.3. Veri Madenciliği Modelleri	10
2.4. Veri Madenciliği Uygulamaları için Temel Adımlar	10
2.5. Temel Veri Madenciliği Problemleri ve Çözüm Yöntemleri	13
3. TEMEL KAVRAMLAR VE MATEMATİKSEL ALTYAPI	17
3.1. Naive Bayes Yöntemi (NB)	17
3.2. Birliktelik Kuralları (Association Rules - AR)	34
3.3. Özelliklerin Önem Sıralaması (Attribute Importance - AI)	36
4. İSTATİSTİKSEL BİLGİLER	38
5. VERİNİN HAZIRLANMASI	75
6. UYGULAMA VE SONUÇLAR	79
6.1. Hatalı Bir Naive Bayes Uygulaması	79
6.2. Sınıflara Göre Mezuniyet Başarısının Tahmini	81
6.2.1. Birinci Sınıf Verilerine Göre Mezuniyet Başarısı Tahmini	81
6.2.1.1. Veri Tablosunun Hazırlanması	83
6.2.1.2. Model İçin Tablo Oluşturma	83
6.2.1.3. Model Oluşturma	87
6.2.1.4. Modelde Etkin Tahmin Edicilerin Belirlenmesi (AI)	89

6.2.1.5. Modelin Testi	92
6.2.1.6. Modelin Uygulanması	94
6.2.1.7. Sonuçlar ve Tartışma	95
6.2.2. İkinci Sınıf Verilerine Göre Mezuniyet Başarısı Tahmini	97
6.2.2.1. Veri Tablosunun Hazırlanması ve Model Oluşturma	98
6.2.2.2. Modelde Etkin Tahmin Edicilerin Belirlenmesi (AI)	99
6.2.2.3. Modelin Testi	102
6.2.2.4. Modelin Uygulanması	103
6.2.2.5. Sonuçlar ve Tartışma	104
6.2.3. Üçüncü Sınıf Verilerine Göre Mezuniyet Başarısı Tahmini	105
6.2.3.1. Veri Tablosunun Hazırlanması ve Model Oluşturma	107
6.2.3.2. Modelde Etkin Tahmin Edicilerin Belirlenmesi (AI)	107
6.2.3.3. Modelin Testi	110
6.2.3.4. Modelin Uygulanması	111
6.2.3.5. Sonuçlar ve Tartışma	112
6.3. Birliktelik Kuralları Uygulaması	113
6.3.1. Veri Tablosunun Hazırlanması	114
6.3.2. Model Oluşturma	115
6.3.3. Birliktelik Kuralları	117
6.3.4. Sonuçlar ve Tartışma	121
KAYNAKLAR	123
EKLER	124

ÖZET

Bu çalışmada İstanbul Teknik Üniversitesi Matematik Mühendisliği Bölümü öğrencileri üzerine bir veri madenciliği uygulaması yapılmıştır. Bu uygulama ile mezun Matematik Mühendisliği öğrencilerinin bölüm dersleri başarı durumları arasındaki ilişkilerin çıkarılması ve derslerden aldıkları harf notları, akademik başarı durumları ile mezuniyet başarı durumlarının tahmin edilmesi amaçlanmaktadır.

Bu amaçla İstanbul Teknik Üniversitesi Öğrenci İşleri'nden 2002-2003 Güz Dönemi'nden 2012-2013 Güz Dönemi'ne kadar Matematik Mühendisliği Bölümü'nde lisans eğitime başlayan öğrencilerin not döküm ve ÖSYM verileri sağlanmış, veriler düzenlenerek tablolar oluşturulmuştur. Mezuniyet başarı durumlarını tahmin etmek için sınıflandırma modeli algoritmalarından Naive Bayes ve bölüm dersleri başarı durumları arasındaki ilişkilerin çıkarılması için Birliktelik Kuralları Apriori Algoritması seçilmiştir.

Veri madenciliği uygulamasını gerçekleştirmek için Oracle Veri Tabanı'nın 11g R2 sürümü ve Oracle SQL Developer 3.2.2 yüklenmiş, SQL Developer'da Oracle Data Miner ile çalışabilmesi için Oracle Data Miner kullanıcısı tanımlanmıştır. Daha sonra oluşturulan tablolar veri tabanına aktarılmış ve gerekli modeller oluşturulmuştur. Bu modeller yardımıyla tatmin edici doğrulukta tahminler elde edilmeye çalışılmış ve sonuçlar yorumlanmıştır.

1. GİRİŞ

Bu bitirme projesi kapsamında; İstanbul Teknik Üniversitesi Öğrenci İşleri Daire Başkanlığı'ndan elde edilen Matematik Mühendisliği öğrenci verileri kullanılarak mezun öğrencilerin mezuniyet başarı durumları ve bölüm dersleri başarı durumları arasındaki ilişkilerin çıkarılması üzerine iki ayrı model oluşturulmuştur. Ham Matematik Mühendisliği öğrenci verilerinden mezun öğrencilerin bilgileri çıkarılmış, modellerde kullanılmaya uygun hale getirilmiştir.

Mezun öğrencilerin mezuniyet başarı durumları tahmini üzerine öğrencilerin birinci, ikinci, üçüncü sınıf bilgileri ile ayrı ayrı modeller kurulmuş, modellerin sonucunda %90'ın üzerinde doğruluk ile tahminler gerçekleştirilmiştir. Bölüm dersleri başarıları için çıkarılmış birliktelik kurallarından %70 ve üzeri güvenilirliğe sahip olanlara yer verilmiştir. Ayrıca modellerin oluşturulmasında kullanılan mezun öğrenciler ve dersler ile ilgili istatistiksel bilgilere de bu çalışma kapsamında yer verilmiştir.

Bölüm dersleri başarı durumları arasındaki ilişkileri ve öğrencilerin mezuniyet başarılarını belirlerken uygulanan veri hazırlanması, model oluşturulması, modelin test edilmesi ve uygulanması aşamaları bu tez kapsamında ifade edilmektedir.

1.1 Açıklamalar

Bu bölümde tez kapsamında kullanılan terimlerin açıklamalarına yer verilmiştir [1].

Tarihçe: Öğrencinin öğrenim hayatı boyunca aldığı derslere ait not döküm bilgileridir. Derslere ait “Dönem”, “CRN”, “Ders Kodu”, “Ders Adı”, “Harf Notu”, “Tekrar Durumu” bilgilerini içerir.

Transfer: Öğrencinin öğrenim hayatı boyunca İTÜ dışından aldığı derslere ait not döküm bilgileridir. Tarihçe ile aynı bilgileri içerir.

Dönem: Dersin alındığı dönemi belirtir. Örneğin, “2003X0” ifadesi 2002-2003 Akademik Yılı'nı belirtirken, ifadedeki “X”; “1”, “2” ve “3” değerleri için sırasıyla dönemin “Güz Dönemi”, “Bahar Dönemi” ve “Yaz Dönemi” olduğunu belirtmektedir.

CRN: Dersin sınıfına ait numaraadır.

Ders Kodu: Dersin kısaltılmıř adıdır.

Ders Adı: Dersin adını belirtir.

Harf Notu: Dnem sonu ders bařarı durumunun harf notu karřılıđını belirtir.

Tekrar Durumu: Dersin ortalamaya katılıp katılmadıđını belirtir. “null”: Dersin notu ortalamaya dahil edilir, “E”: Dersin notu ortalamaya dahil edilmez, “I”: Tekrar edilmiř derstir ve dersin notu ortalamaya dahil edilir řeklinde bilgi ierir.

đrenci No: đrenci ID’sidir.

Kabul Tr: İT Matematik Mhendisliđi Blm’ne kabuln sađlayan durumdur. “Standart SS”: SYM’nin yapmıř olduđu sınav ile blme kabul edilen đrenciyi belirtir. “Standart YS”: YS sınavıyla blme kabul edilen đrenciyi belirtir. “Ek kontenjan SS”: SYM’nin yapmıř olduđu sınav ile blmde kalan boř kontenjandan yararlanıp blme kabul edilen đrenciyi belirtir. “niversite İinden Yatay Geiř”: niversite iinden yatay geiř ile blme kabul edilen đrenciyi belirtir. “niversite Dıřından Yatay Geiř”: niversite dıřından yatay geiř ile blme kabul edilen đrenciyi belirtir. “5316 No’lu Af Kanunu”: 5316 No’lu Af Kanunu ile blme kabul edilen đrenciyi belirtir. “6111 No’lu Af Kanunu”: 6111 No’lu Af Kanunu ile blme kabul edilen đrenciyi belirtir.

İkinci Anadal: ift Anadal Programı đrencisinin ikinci anadalını belirtir.

Dnem GPA: Sadece dnem dersleri dahil edilerek hesaplanan ortalamayı belirtir.

Toplam GPA: Tm dersleri dahil edilerek hesaplanan ortalamayı belirtir.

Akademik Durum: Ortalamaya gre belirlenmiř durumdur. “İyi Durum”: İyi durumda olunduđunu belirtir. “Gzetim”: Gzetim durumunda olunduđunu belirtir.

Mezuniyet Tarihi: Mezun olunan tarih bilgisidir.

Mezuniyet Dnemi: Mezun olunan tarihin denk geldiđi dnemdir.

SYM Puanı: SYM’nin yaptıđı sınavda alınan puanı belirtir.

SYM Bařarı Sırası: SYM’nin yaptıđı sınavdaki bařarı sırasını belirtir.

Giriş Türü: Öğrencinin Matematik Mühendisliği bölümüne hangi kontenjan dahilinde girdiğini belirtir. “Genel”: Genel kontenjan ile girildiğini belirtir. “Okul Birinciliği”: Okul birinciliği kontenjanıyla girildiğini belirtir.

Irregular (düzensiz) öğrenci: Bir yarıyıl hazırlık sınıfına devam edip, öğrenimlerine ikinci yarıyıl da başlayan öğrencileri belirtir.

Regular (düzenli) öğrenci: Öğrenimlerine güz yarıyılında başlayan öğrencileri belirtir.

1.2 Öğretim Süresi ve Öğretim Planları

Lisans öğretimi İngilizce hazırlık öğretimindeki süre hariç en fazla yedi yıldır. Bu süre sonunda mezun olamayan öğrencilerin, ders ve sınavlara katılma ile bitirme tasarımı hariç, öğrencilere tanınan diğer haklardan yararlanılmadan öğrencilik statüleri devam eder.

Üniversite fakülte ve bölümlerindeki öğretim, öğretim planına [2] göre yapılır. Mezun Matematik Mühendisliği Öğrencileri’nin sorumlu oldukları öğretim planlarının dönemlere göre değişiklik gösterdiği Tablo 1.1’den görülmektedir.

Tablo1.1: Dönemlere göre ders planları

Ders Planı	Uygulandığı Dönemler
Ders Planı 1	2000 – 2001 / Güz ile 2004 – 2005 / Güz Dönemleri Arası
Ders Planı 2	2004 – 2005 / Güz ile 2007 – 2008 / Güz Dönemleri Arası
Ders Planı 3	2007 – 2008 / Güz ile 2009 – 2010 / Güz Dönemleri Arası
Ders Planı 4	2009 – 2010 / Güz ile 2012 – 2013 / Güz Dönemleri Arası

Ders planları, yarıyıllara göre Tablo1.2, Tablo1.3, Tablo1.4, Tablo1.5, Tablo1.6, Tablo1.7, Tablo1.8, Tablo1.9’da verilmektedir. Ders planlarında zorunlu dersler ders koduyla, bölüm dışı seçmeli dersler, ITB ve bölüm seçmeli dersleri ise MT ile belirtilmektedir.

Tablo1.2: Ders planlarına göre 1. yarıyıl dersleri ve kredileri

	MAT 111	MAT 175	MAT 141	FIZ 101E	FIZ 101EL	KIM 101E	KIM 101EL	BİL 101E	English Course I
Ders Planı 1	5	3	4	3	1	3	1	1.5	3
Ders Planı 2	5	3	4	3	1	3	1	1.5	3
Ders Planı 3	5	3	3	3	1	3	1	1.5	3
Ders Planı 4	5	3	3	3	1	3	1	1.5	3

Tablo1.3: Ders planlarına göre 2. yarıyıl dersleri ve kredileri

	MAT 112	MAT 142	MAT 122	FIZ 106	FIZ 106L	Intro to Sci & Eng Comp	BİL 104E	English Course II
Ders Planı 1	5		3	3	1	3		3
Ders Planı 2	5		3	3	1	3		3
Ders Planı 3	5	3		3	1	3		3
Ders Planı 4	5	3		3	1		3	3

Tablo1.4: Ders planlarına göre 3. yarıyıl dersleri ve kredileri

	MAT 213	MAT 322E	MAT 221	MUH 211	DNK 201E	MUK 201E	TUR 101
Ders Planı 1	4		3	3	4	4	2
Ders Planı 2	4	4	3	3	3		2
Ders Planı 3	4	4	3	3	3		2
Ders Planı 4	4	4	3	3	3		2

Tablo1.5: Ders planlarına göre 4. yarıyıl dersleri ve kredileri

	MAT214	MAT232	MAT242	MUH212	ING201	TUR102
Ders Planı 1	3	4	3	3	3	2
Ders Planı 2	3	4	4	3	3	2
Ders Planı 3	3	4	4	3	3	2
Ders Planı 4	3	4	4	3	3	2

Tablo1.6: Ders planlarına göre 5. yarıyıl dersleri ve kredileri

	MAT331	MAT341E	MUH311	MUH321	MT	ITB
Ders Planı 1	4	4	3	4	3	3
Ders Planı 2	4	4	3	4	3	3
Ders Planı 3	4	4	3	4	3	3
Ders Planı 4	4	4	3	4	3	3

Tablo1.7: Ders planlarına göre 6. yarıyıl dersleri ve kredileri

	MAT 355	MAT 332	MAT 322E	MAT 342E	MUH 312	MT	ITB
Ders Planı 1		4	4	3	3	3	3
Ders Planı 2	3	4		3	3	3	3
Ders Planı 3	3	4		3	3	3	3
Ders Planı 4	3	4		3	3	3	3

Tablo1.8: Ders planlarına göre 7. yarıyıl dersleri ve kredileri

	MT1	MT2	MT3	MT4	ITB	ATA101
Ders Planı 1	3	3	3	3	3	2
Ders Planı 2	3	3	3	3	3	2
Ders Planı 3	3	3	3	3	3	2
Ders Planı 4	3	3	3	3	3	2

Tablo1.9: Ders planlarına göre 8. yarıyıl dersleri ve kredileri

	MT5	MT6	ATA102	EKO201	MAT492
Ders Planı 1	3	3	2	3	3
Ders Planı 2	3	3	2	3	3
Ders Planı 3	3	3	2	3	3
Ders Planı 4	3	3	2	3	3

Öğrenci verileri içinde en fazla Ders Planı 3'e göre ders almış mezun standart öğrenci olduğu için **öğrenci ders verileri Ders Planı 3'e göre düzenlenmiştir** ve derslerin Ders Planı 3'teki kredileri baz alınmıştır. Yani tüm öğrencilerin 39 zorunlu ders, 8 MT, 3 ITB dersi ile ve minimum 151.5 kredi ile mezun olmaları beklenmelidir.

1.3 Notlar ve İşaretler

Öğrencinin bir derste başarı, öğrencinin dönem içerisinde gösterdiği başarının ve dönem sonu sınavlarında aldığı notun sınıfın genel durumu ile değerlendirilmesiyle bulunur ve başarı harf notu karşılığı verilir. Harf notu **AA, BA, BB, CB, CC, DC, DD, FF, VF** olabilir. Harf notlarının yansıttığı başarı durumu ve not karşılığı Tablo1.10'da verilmiştir.

Tablo1.10: Harf notlarının başarı ve not karşılıkları

HARF NOTU	BAŞARI DURUMU	BAŞARI NOTU
AA	Dersi Başardı	4.00
BA	Dersi Başardı	3.50
BB	Dersi Başardı	3.00
CB	Dersi Başardı	2.50
CC	Dersi Başardı	2.00
DC	Dersi Şartlı Başardı	1.50
DD	Dersi Şartlı Başardı	1.00
FF	Dersten Başarısız	0.00
VF	Dersten Başarısız	0.00

Başarı harf notları dışında verilen harfler Tablo1.11’de verilmiştir. Bu harflerin not karşılığı yoktur. Öğrencinin not dökümünde yer alır fakat öğrencinin ortalamasına etki etmez.

Tablo1.11: Başarı harf notları dışındaki harfler ve not karşılıkları

HARF	KARŞILIK
BL	Başarılı
BZ	Başarısız
T	Terk
E	Erteleme
M	Muaf

1.4 Ağırlıklı Genel Not Ortalaması ve Gözetim Durumu

Ağırlıklı genel not ortalaması (AGNO); öğrencinin her bir dersi için yapılan, dersin başarı notunun o dersin kredi karşılığıyla çarpılması işleminden elde edilen sayıların toplamının, krediler toplamına bölünmesi ile bulunur.

Öğrencilerin akademik durumları her dönem sonunda ağırlıklı genel not ortalamalarına göre belirlenir. Ağırlıklı genel not ortalamasının eşik değeri her sınıf için ayrı belirlenmiştir. Bu değer altında kalan öğrencinin akademik durumu “Gözetim Durumu”, eşit ve üzerinde olanların ise “İyi Durum” olarak belirlenir. Sınıflar ve ağırlıklı genel not ortalaması eşik değeri Tablo1.12’de verilmiştir.

Tablo1.12: Sınıflar ve ağırlıklı genel not ortalaması

SINIF	AGNO
1	1.80
2	1.80
3	1.90
4	2.00

Gözetim durumundaki öğrenciler, bir yarıyılıda en fazla 15 kredilik ders alabilir. Gözetim durumunda iken ağırlıklı genel not ortalamalarını belirtilen düzeylerin üzerine çıkartan öğrenciler iyi duruma geçerek, normal statüde öğrenimlerine devam ederler.

Ağırlıklı genel not ortalaması, daha önce alınıp başarılan derslerin tekrar alınmasıyla yükseltilebilir. Bu dersler için son alınan harf notları geçerlidir. Öğrencinin, öğrenim hayatı boyunca aldığı tüm dersler için bütün harf notları ve işaretleri not döküm belgesinde yer almaktadır.

Öğretim planında yer alan zorunlu ve seçmeli dersleri başaran ve ağırlıklı genel not ortalamasını en az 2.00'a çıkaran öğrenciler lisans öğrenimlerini bitirirler. Öğrenim sonundaki ağırlıklı genel not ortalaması (mezuniyet ortalaması)'na göre öğrencinin akademik durumu belirlenir.

2. VERİ MADENCİLİĞİ

Günümüzde kullanılan veri tabanı yönetim sistemleri eldeki verilerden sınırlı çıkarımlar yaparken geleneksel çevrimiçi işlem sistemleri (on-line transaction processing systems) de bilgiye hızlı ve güvenli erişimi sağlamaktadır. Fakat her ikisi de eldeki verilerden analizler yapıp anlamlı bilgiler sağlamakta yetersiz kalmışlardır. Verilerin yığınla artması ve anlamlı çıkarımlar elde etme ihtiyacı arttıkça uzmanlar “Knowledge Discovery in Databases” (KDD) veri tabanında bilgi keşfi adı altında çalışmalarına hız kazandırmışlardır. Bu çalışmalar sonucunda da "Data mining" veri madenciliği kavramı doğmuştur. Veri madenciliğinin temel amacı, çok büyük veri tabanlarındaki ya da veri ambarlarındaki veriler arasında bulunan ilişkiler, örüntüler, değişiklikler, sapma ve eğilimler, belirli yapılar gibi bilgilerin matematiksel teoriler ve bilgisayar algoritmaları kombinasyonları ile ortaya çıkartılması ve bunların yorumlanarak değerli bilgilerin elde edilmesidir [3].

2.1 Tanım

İlişkisel veri tabanı sistemleriyle ulaşılan veriler tek başına bir anlam ifade etmezken, veri madenciliği teknolojisi bu verilerden anlamlı bilgi üretmede öncü rol oynamaktadır.

Veri madenciliği; matematiksel yöntemler yardımıyla, biriken veri yığınları içerisinde bulunan verilerin birbirleriyle ilişkisini ortaya çıkartmak için yapılan analiz ve kurulan modeller sonucunda elde edilecek bilgi keşfi sürecidir.

Veri madenciliğinin, disiplinler arası bir teknoloji olarak dört ana başlıktan oluştuğu kabul edilmektedir. Bunlar sınıflama, kategori etme, tahmin etme ve görüntülemedir. Bu dört temel başlık dışında istatistik, makine bilgisi, veri tabanları ve yüksek performanslı işlem gibi disiplinleri de içerir.

2.2 Kullanım Alanları

Günümüzde veriler çok hızlı bir şekilde toplanabilmekte, depolanabilmekte, işlenebilmekte ve bilgi olarak kurumların hizmetine sunulabilmektedir. Bilgiye hızlı erişim firmaların sürekli yeni stratejiler geliştirip etkili kararlar almalarını sağlayabilmektedir. Bu süreçte araştırmacılar büyük hacimli ve dağınık veri setleri üzerinde firmalara gerekli bilgi keşfini daha hızlı gerçekleştirebilmeleri için veri madenciliği üzerine çalışmalar yapmaktadırlar. Tüm bu çalışmalar doğrultusunda veri madenciliği günümüzde yaygın bir kullanım alanı bulmaktadır. Veri madenciliği, perakende ve pazarlama, bankacılık, sağlık hizmetleri, sigortacılık, tıp, ulaştırma, eğitim, ekonomi, güvenlik, elektronik ticaret alanında yaygın olarak kullanılmaktadır [3].

2.3 Veri Madenciliği Modelleri

IBM tarafından veri işleme operasyonları için iki çeşit model tanımlanmıştır.

2.3.1 Doğrulama modeli

Doğrulama modeli kullanıcıdan bir hipotez olarak testler yapar ve bu hipotezin geçerliliğini araştırır.

2.3.2 Keşif modeli

Sistem bu modelde önemli bilgileri gizli veriden otomatik olarak elde eder. Veri başka hiçbir aracıya ihtiyaç duymadan yaygın olarak kullanılan modeller, genelleştirmeler ile ayıklanır.

2.4 Veri Madenciliği Uygulamaları İçin Temel Adımlar

Veri madenciliği uygulamalarında sırasıyla takip edilmesi gereken temel aşamalar aşağıda sistematik biçimde verilmiştir.

2.4.1 Uygulama alanının ortaya konulması

Bu ilk adımda veri madenciliğinin hangi alan ve hangi amaç için yapılacağı tespit edilir.

2.4.2 Hedef veri grubu seçimi

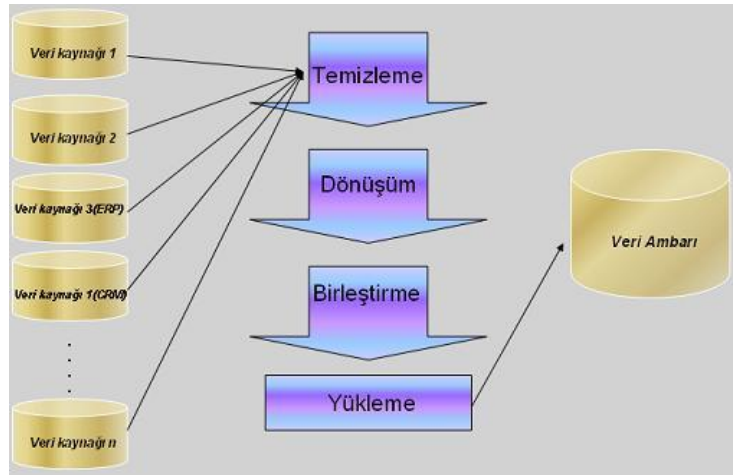
Belirlenen amaç doğrultusunda bazı kriterler belirlenir. Bu kriterler çerçevesinde aynı veya farklı veri tabanlarından veriler toplanarak "target" hedef veri grubu elde edilir.

2.4.3 Model seçimi

Veri madenciliği probleminin seçimi veriler üzerinden belirlenir. (Sınıflandırma, Kümeleme, Birliktelik Kuralları, Şablonların ve İlişkilerin Yorumlanması v.b.)

2.4.4 Ön işleme

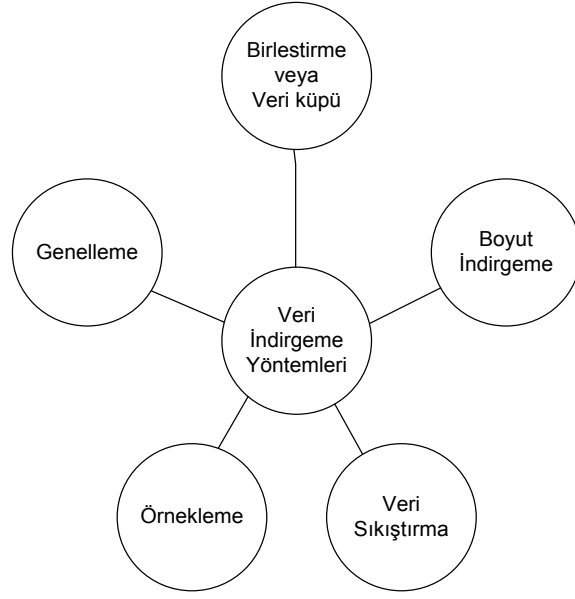
Bu aşamada seçilen veriler ayıklanarak silinir, eksik veri alanları üzerine stratejiler geliştirilir. Veriler tekrardan düzenlenip tutarlı bir hale getirilir. Bu aşamada yapılan işlem veri temizleme ve veri birleştirme olarak bilinen uyumlandırma işlemidir. Veri birleştirme (bütünleştirme), farklı veri tabanlarından ya da kaynaklarından elde edilen verilerin birlikte değerlendirmeye alınabilmesi için farklı türdeki verilerin tek türe dönüştürülebilmesi demektir. Bu süreç Şekil 2.1’de gösterilmektedir.



Şekil 2.1 : Ön işleme süreci

2.4.5 Veri indirgeme

İşlem yapılırken eksik ya da uygun olmayan verilerin oluşturduğu tutarsız verilerle karşılaşılabilir. Bu gibi durumlarda verinin söz konusu sorunlardan arındırılması gerekmektedir. Çözümlemeyen elde edilecek sonuçta bir değişiklik olmuyorsa veri sayısı ya da değişkenlerin sayısında azaltmaya gidilir. Veri indirgeme çeşitli biçimlerde yapılabilir. Bu yöntemler Şekil 2.2’de gösterilmiştir [4].



Şekil 2.2 : Veri indirgeme yöntemleri

Veriyi indirirken bu verileri çok boyutlu veri küpleri biçimine dönüştürmek söz konusu olabilir. Böylece çözümler sadece belirlenen boyutlara göre yapılır. Veriler arasında seçme işlemi uygulanarak da veri tabanından veriler silinip boyut azaltılması yapılmaktadır.

2.4.6 Veri dönüştürme

Değişkenlerin ortalama ve sapmaları birbirlerinden çok farklıysa büyük ortalama ve sapmaya sahip değişkenlerin diğer değişkenler üzerindeki etkisi daha fazla olur. Bu nedenle bir dönüşüm yöntemi uygulanarak değişkenlerin normalleştirilmesi ya da standartlaşması uygun yoldur.

2.4.7 Algoritmanın belirlenmesi

Bu aşamada indirgenmiş veriye ve kullanılacak modele hangi algoritmanın uygulanacağına karar verilir.

2.4.8 Yorumlama ve doğrulama

Uygulama sonucunda elde edilen veriler üzerine yorumlama yapılır, bu yorumların test verileri üzerinden doğrulanması hedeflenir. Doğruluğu onaylanan bu yorumlar gizli bilgiye ulaşıldığını göstermektedir. Elde edilen bu bilgiler çoğu kez grafiklerle desteklenir.

2.5 Temel Veri Madenciliği Problemleri ve Çözüm Yöntemleri

Veri madenciliği uygulaması gerektiren problemlerde, farklı veri madenciliği algoritmaları ile çözüme ulaşılmaktadır. Veri madenciliği görevleri iki başlık altında toplanmaktadır:

- Eldeki verinin genel özelliklerinin belirlenmesidir.
- Kestirimci/Tahmin edici veri madenciliği görevleri, ulaşılabilir veri üzerinde tahminler aracılığıyla çıkarımlar elde etmek olarak tanımlanmıştır.

Veri madenciliği algoritmaları aşağıda açıklanmaktadır.

2.5.1 Karakterize etme (Characterization)

Veri karakterizasyonu hedef sınıfındaki verilerin seçilmesi, bu verilerin genel özelliklerine göre karakteristik kuralların oluşturulması olayıdır.

Örnek: Perakende sektöründe faaliyet gösteren, uluslararası ABC şirketinin binlerce kurumsal müşterisi olsun. ABC şirketinin pazarlama biriminde, büyük kurumsal müşterilere yönelik kampanyalar için her yıl düzenli olarak bu şirketten 10 milyon TL ve üstü alım yapan kurumsal müşteriler hedeflenmektedir. Veri tabanından hedef grup belirlenerek genelleme yapılır ve genel kurallar oluşturulur.

2.5.2 Ayrıştırma (Discrimination)

Belirlenen hedef sınıfa karşıt olan sınıf elemanlarının özellikleri arasında karşılaştırma yapılmasını sağlayan algoritmadır. Karakterize etme metodundan farkı mukayese yöntemini kullanmasıdır.

Örnek: ABC kurumsal müşterilerinden her yıl 10 milyon TL ve üstü alışveriş yapan fakat geri ödeme konusunda riskli olan müşteri grubunun belirlenmesi

2.5.3 Sınıflandırma (Classification)

Sınıflandırma, veri tabanlarındaki gizli örüntüleri ortaya çıkarabilmek için veri madenciliği uygulamalarında sıkça kullanılan bir yöntemdir. Verilerin sınıflandırılması için belirli bir süreç izlenir. Öncelikle var olan verinin bir kısmı model oluşturmada kullanılarak sınıflandırma kurallarının oluşturulması sağlanır. Bu kurallar kullanılarak veriler sınıflandırılır. Bu veriler sınıflandırıldıktan sonra eklenecek yeni veriler bu sınıflardan karakteristik olarak uygun olan kısma atanır.

Sınıflandırma problemleri için “Oracle Data Miner” (ODM)’ in uyumlu olduğu çözüm yöntemleri Naive Bayes (NB), Karar Destek Vektörleri (SVM), Karar Ağaçları [3].

Örnek: XYZ şirketi müşterilerinin alım durumlarını göz önünde bulundurarak, alım gücüne göre “Yüksek”, “Orta”, “Düşük” şeklinde sınıflandırır. Müşterilerinin risk durumlarını sınıflandırmak için de “Risksiz”, “Riskli”, “Çok Riskli” şeklinde etiketlerle sınıflandırılabilir.

2.5.4 Tahmin etme (Prediction)

Kayıt altında tutulan geçmiş verilerin analizi sonucu elde edilen bilgiler gelecekte karşılaşılabilecek aynı tarz bir durum için tahmin niteliği taşıyacaktır. Örneğin ABC şirketi geçen yılın satışlarını bölge bazlı sınıflandırmış ve bu sene için bir eğilim analizi yaparak her bölgede oluşacak talebi tahmin etmiştir. Bu tür problemler için ODM’nin kullandığı regresyon analizi yöntemi SVM’dir.

2.5.5 Birliktelik kuralları (Association rules)

Birliktelik kuralları gerek birbirini izleyen gerekse de eş zamanlı durumlarda araştırma yaparak, bu durumlar arasındaki ilişkilerin tanımlanmasında kullanılır. Bu modelin yaygın olarak Market Sepet Analizi uygulamalarında kullanıldığı bilinmektedir. Örneğin bir süpermarkette X ürününden alan müşterilerin büyük bir kısmı Y ürününden de almıştır. Birliktelik kuralı ile bu durum ortaya çıkarılarak, süpermarketin X ve Y ürününü aynı veya yakın raflara koyması sağlanır. ODM bu problem sınıfı için de Birliktelik Kuralları modelini kullanmaktadır.

2.5.6 Kümeleme (Clustering)

Yapı olarak sınıflandırmaya benzeyen kümeleme metodunda birbirine benzeyen veri grupları aynı tarafta toplanarak kümelenmesi sağlanır. Sınıflandırma metodunda sınıfların kuralları, sınırları ve çerçevesi belli ve veriler bu kriterlere göre sınıflara atanırken kümeleme metodunda sınıflar arası bir yapı mevcut olup, benzer özellikte olan verilerle yeni gruplar oluşturmak asıl hedeftir. Verilerin kendi aralarındaki benzerliklerinin göz önüne alınarak gruplandırılması yöntemin pek çok alanda uygulanabilmesini sağlamıştır. Örneğin, pazarlama araştırmalarında, desen tanımlama, resim işleme ve uzaysal harita verilerinin analizinde kullanılmaktadır.

Tüm bu uygulama alanlarında kullanılması, ODM'nin desteklediği “K-means” ve “O-Cluster” kümeleme yöntemleri ile mümkün kılınmıştır.

2.5.7 Aykırı değer analizi (Outlier analysis)

İstisnalar veya sürpriz olarak tespit edilen aykırı veriler, bir sınıf veya kümelemeye tabii tutulamayan veri tipleridir. Aykırı değerler bazı uygulamalarda atılması gereken değerler olarak düşünülürken bazı durumlarda ise çok önemli bilgiler olarak değerlendirilebilmektedir.

Örneğin bir markette müşterilerin hep aynı ürünü iade etmesi bu metodun araştırma konusu içine girer. ODM temizleme, eksik değer, aykırı değer analizi gibi birçok yöntemi veri hazırlama aşaması içine almakta ve desteklemektedir.

2.5.8 Zaman serileri (Time series)

Yapılan veri madenciliği uygulamalarında kullanılan veriler çoğunlukla statik değildir ve zamana bağlı olarak değişmektedir. Bu metot ile bir veya daha fazla niteliğin belirli bir zaman aralığında, eğilimindeki değişim ve sapma durumlarını inceler. Belirlenen zaman aralığında ölçülebilir ve tahmin edilen/beklenen değerleri karşılaştırmalı olarak inceler ve sapmaları tespit eder.

Örneğin ABC şirketinin Ocak-Haziran 2009 dönemi için önceki yılın satış miktarları göz önünde tutularak bir hedef ortaya konulmuştur. 2008 ve 2009 değerleri karşılaştırmalı olarak incelenerek sapma miktarı belirlenir. ODM her ne kadar çeşitli histogramlarla kullanıcıya görsel destek sağlasa da tam anlamıyla bu tür problemleri desteklememektedir.

2.5.9 Veri görüntüleme (Visualization)

Bu metot çok boyutlu özelliğe sahip verilerin içerisindeki karmaşık bağlantıların/bağıntıların görsel olarak yorumlanabilme imkanını sağlar. Verilerin birbirleriyle olan ilişkilerini grafik araçları görsel ya da grafiksel olarak sunar. ODM zaman serilerinde olduğu gibi histogramlarla bu metodu desteklemektedir.

2.5.10 Yapay sinir ağları (Artificial neural networks)

Yapay sinir ağları insan beyninden esinlenilerek geliştirilmiş, ağırlıklı bağlantılar aracılığıyla birbirine bağlanan işlem elemanlarından oluşan paralel ve dağıtılmış bilgi işleme yapılarıdır. Yapay sinir ağları öğrenme yoluyla yeni bilgiler türetebilme ve

keşfedebilme gibi yetenekleri hiçbir yardım almadan otomatik olarak gerçekleştirebilmek için geliştirilmiştir. Yapay sinir ağlarının temel işlevleri arasında veri birleştirme, karakterize etme, sınıflandırma, kümeleme ve tahmin etme gibi veri madenciliğinde de kullanılan metotlar mevcuttur. Yüz ve plaka tanıma sistemleri gibi teknolojiler yapay sinir ağları kullanılarak geliştirilen teknolojilerdendir.

2.5.11 Genetik algoritmalar (Genetic algorithms)

Genetik algoritmalar doğada gözlemlenen evrimsel sürece benzeyen, genetik kombinasyon, mutasyon ve doğal seçim ilkelerine dayanan bir arama ve optimizasyon yöntemidir. Genetik algoritmalar parametre ve sistem tanılama, kontrol sistemleri, robot uygulamaları, görüntü ve ses tanıma, mühendislik tasarımları, yapay zeka uygulamaları, fonksiyonel ve kombinasyonel en iyileme problemleri, ağ tasarım problemleri, yol bulma problemleri, sosyal ve ekonomik planlama problemleri için diğer eniyileme yöntemlerine kıyasla daha başarılı sonuçlar vermektedir.

2.5.12 Karar ağaçları (Decision trees)

Ağaç yapıları esas itibarıyla kural çıkarma algoritmaları olup, veri kümelerinin sınıflanması için “if-then” tipinde kullanıcının rahatlıkla anlayabileceği kurallar inşa edilmesinde kullanılırlar. Karar ağaçlarında veri kümesini sınıflamak için “Classification and Regression Trees (CART)” ve “Chi Square Automatic Interaction Detection (CHAID)” şeklinde iki yöntem kullanılmaktadır.

2.5.13 Kural çıkarma (Rules induction)

İstatistiksel öneme sahip yararlı “if-else” kurallarının ortaya çıkarılması problemlerini inceler.

3. TEMEL KAVRAMLAR VE MATEMATİKSEL ALTYAPI

3.1 Naive Bayes Yöntemi (NB)

Bu bölümde veri kümesine uygulamak üzere seçilen sınıflandırma modelinin kullandığı Naive Bayes yönteminin matematiksel altyapısından bahsedilmektedir.

Sınıflandırma problemlerinin çözümünde kullanılan Naive Bayes yöntemi temel olarak olasılık teorisini kullanmaktadır. Bu bölümde önce yöntemin teorisi, sonrasında da yöntemle ilgili küçük örnekler verilmektedir.

3.1.1 Temel kavramlar

Olasılık: Bir olayın olabilirliğinin ölçüsüdür, $[0-1]$ arasında değer alabilir, $P(A)$ ile gösterilir ve

$P(A) = 1$, A olayının mutlaka gerçekleşeceğini

$P(A) = 0$, A olayının gerçekleşmesinin mümkün olmadığını ifade eder.

Vektör: Burada kullanacağımız anlamıyla bir satır vektör $\mathbf{x} = \{x_1, x_2, x_3, \dots, x_{m-1}, x_m\}$ şeklinde m elemanı ile belirlenen ve i . elemanı x_i ile verilen bir büyüklüktür.

Veri madenciliği uygulanacak olan veri kümesi aşağıda gösterilen tablonun formatındadır. Tabloda her satır (her kayıt) bir vektör ($\mathbf{x}_i$) olarak düşünülür, $\mathbf{x}_i$ vektörünün j . elemanı i . kaydın A_j sütunundaki değerine karşı gelir. Son sütun (B) yani $\mathbf{y}$ sütun vektörü, veri madenciliği ile tahmin edilmek istenen hedef özelliktir. Dolayısıyla n kayıt ve $(m+1)$ sütundan oluşan bir tabloda her biri m boyutlu n tane belirleyici $\mathbf{x}_i$ vektörü ve bir tane hedef sütun (B), yani $\mathbf{y}$ vektörü vardır.

Tahmin edici sütunlar (özellikler)							Sonuç, hedef sütun
Sütunlar	A ₁	A ₂	A ₃		A _{m-1}	A _m	
1. kayıt X ₁							
2. kayıt X ₂							
3. kayıt X ₃							
⋮							
(n-1). kayıt X _{n-1}							
n. kayıt X _n							

Şekil 3.1 : Bir veri kaydı örneği

3.1.2 Teori

Naive Bayes yöntemi ile sınıflandırma koşullu olasılık hesabına dayanmaktadır. Şekil 3.1’de görüldüğü üzere tüm değerleri belirli geçmiş bir veri kümesinde, B yani sonuç sütunu, diğer $A_i, (i=1,...,m)$ sütunlarına bağlı kabul edilerek, $P(B=b_j | A_i=a_{ik}, ... (i=1,...,m))$, olasılıkları hesaplanır, burada $j=1,...,s$ ve $k=1,...,m_i$ dir. Bu ifade ile her biri m_i tane farklı gruptan oluşan A_i sütunları a_{ik} değerlerini aldıklarında, bu A_i sütunlarına bağlı olarak, B sütununda bulunan s tane farklı grubun b_j değerlerinden her birini alma olasılıkları hesaplanmaktadır. Geçmiş veri kümesi yardımıyla hesaplanan bu olasılıklar, yeni gelecek verinin hangi gruba dahil edileceğinin yani B sütununun tahmininde kullanılacaktır.

Konuyu anlaşılır kılmak için tahmin edici sütun önce bir tane A_1 , sonra iki tane A_1, A_2 alınarak, B sütununun bunlara bağlı olasılıkları hesaplanarak problem basitleştirilmiş daha sonra ise m sütun alınarak problem genelleştirilmiştir.

Burada öncelikle koşullu olasılık kavramının açıklanması gerekmektedir. A ve B iki olay olmak üzere, bu olayların olma olasılıkları $P(A)$ ve $P(B)$ ile verilir. Eğer A ve

B olaylarının gerçekleşmesi birbirine bağlı değilse, bu iki olayın birlikte olma olasılığı

$$P(A, B) = P(A) \times P(B) \quad (3.1)$$

ile verilir. Örneğin A olayı, o gün havanın yağmurlu olması ve B olayı ise atılan bir madeni paranın yazı gelme olasılığı ise, bu iki olay birbirinden bağımsızdır ve bu iki olayın birlikte olma olasılıkları her bir olayın olma olasılıklarının çarpımına eşittir.

Eğer A ve B olayları birbirine bağlı ise, bu iki olayın birlikte olma olasılıkları; A 'nın olma olasılığı ile A 'dan sonra B 'nin olma olasılığının çarpımı ile yani

$$P(A, B) = P(A) P(B|A) \quad (3.2)$$

veya B 'nin olma olasılığı ile B 'den sonra A 'nın olma olasılığının çarpımı ile yani

$$P(A, B) = P(B) P(A|B) \quad (3.3)$$

ile verilir. Dolayısıyla (3.2) ve (3.3) denklemleri birbirine eşitlenerek, A olayından sonra B olayının olma olasılığı

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} \quad (3.4)$$

ile verilir. Örneğin A olayı havanın yağmurlu olması, B olayı ise Ali'nin balığa çıkma olayı ise, B olayının A olayına bağlı olduğu açıktır ve A olayından sonra B olayının olma olasılığı yani hava yağmurlu iken Ali'nin balığa çıkma olayı (3.4) ifadesiyle hesaplanır.

Bir olayın olması ve olmaması olasılıkları toplamı $P(B) + P(B^\perp) = 1$ 'dir. Burada " $^\perp$ " üst indis B olayının değilini göstermektedir. Dolayısıyla Ali hava yağmurlu iken balığa çıktığı gibi, yağmur yağmazken de balığa çıkabilir, yani bir B olayına bağlı olarak A olayının olma olasılığı

$$P(A) = P(A, B) + P(A, B^\perp) = P(B)P(A|B) + P(B^\perp)P(A|B^\perp) \quad (3.5)$$

şeklinde verilir. Bu ifade, (3.4)'te kullanılırsa,

$$P(B|A) = \frac{P(B)P(A|B)}{P(B)P(A|B) + P(B^\perp)P(A|B^\perp)} \quad (3.6)$$

elde edilir. Eğer A ve B olayları farklı değerler alabiliyorsa, örneğin Ali'nin balığa çıkması (b_1), işe gitmesi (b_2), spor yapması (b_3) gibi üç farklı B olayı varsa bu

durumda $P(B=b_1)+P(B=b_2)+P(B=b_3)=1$ dir. (3.5) ifadesine benzer bir şekilde bu kez A olayı r tane ayrık a_k ve B olayı s tane ayrık b_j değeri alıyorsa;

$$P(A=a_k)=\sum_{j=1}^s P((A=a_k), (B=b_j))=\sum_{j=1}^s P(B=b_j)P((A=a_k)|(B=b_j)) \quad (3.7)$$

elde edilir. (3.7) ifadesi (3.4)'te yerine yazıldığında ise,

$$P((B=b_j)|(A=a_k))=\frac{P(B=b_j)P((A=a_k)|(B=b_j))}{\sum_{k=1}^s P(B=b_k)P(A|(B=b_k))} \quad (3.8)$$

elde edilir. (3.8) ifadesinin A ve B olaylarının ikiden fazla değer alabildikleri durum için (3.6) ifadesinin genelleştirilmiş hali olduğu açıktır. Bu ifade Şekil.3.1'de verilen tabloda B sonuç sütununu tahmin edici tek bir A_1 sütunu olması halinde B sütununun alabileceği değerlerin olasılıklarının hesaplanmasında kullanılır. Ancak gerçekte sadece biri tahmin edici, diğeri hedef sütun olmak üzere iki sütun olması değil, hedef sütunu tahmin edici birçok sütun bulunması beklenir.

Bu nedenle (3.8) ifadesinde A gibi sadece bir tahmin edici sütun yerine m tane A_i sütunu olduğunu ve bunların her birinin r_i tane bağımsız değer alabildiği yani örneğin A_1 sütunu $r_1=5$, A_2 sütunu $r_2=3$ farklı değer alabildiğini varsayalım. Bu durumda (3.8) ifadesinde A yerine $A_1, A_2, \dots, A_m$ gibi m tane olay alınır;

$$P(B=b_j | A_1=a_{1j_1}, A_2=a_{2j_2}, \dots, A_m=a_{mj_m}) = \frac{P(B=b_j)P(A_1=a_{1j_1}, A_2=a_{2j_2}, \dots, A_m=a_{mj_m} | B=b_j)}{\sum_{k=1}^s P(B=b_k)P(A_1=a_{1j_1}, A_2=a_{2j_2}, \dots, A_m=a_{mj_m} | B=b_k)} \quad (3.9)$$

ifadesi elde edilir. Tahmin edici her sütunun yani her A_i olayının birbirinden bağımsız olduğu kabulü yapılırsa, sonuç olarak

$$P(B=b_k | A_1=a_{1j_1}, A_2=a_{2j_2}, \dots, A_m=a_{mj_m}) = \frac{P(B=b_k) \times \prod_{i=1}^m P(A_i=a_{ij_i} | B=b_k)}{\sum_{\forall r|b_r \in B} \left(P(B=b_r) \times \prod_{i=1}^m P(A_i=a_{ij_i} | B=b_r) \right)} \quad (3.10)$$

ifadesi elde edilir. Burada $j_i = 1, \dots, m_i$ ve $k = 1, \dots, s$ için bu olasılık değerleri hesaplanmalıdır, ayrıca $\forall r | b_r \in B$ terimi hedef sütunun alabileceği tüm farklı değerler üzerinde toplam alınacağını ifade etmektedir [5].

3.1.3 Örnekler

Örnek 1) Tek boyut için: Yapılan bir anket sonucunda 100 deneğin maaşları “az”, “orta” ve “yüksek” olarak gruplanmış ve hangi gazeteyi okudukları ise ikinci bir sütunda Tablo 3.1a’daki gibi belirtilmiş olsun.

Oracle Data Miner (ODM) verilerin olasılık hesaplarını arka planda otomatik olarak işleyip kullanıcıya sadece sonucu bildirir. Anket sonuçları kullanılarak ODM’nin arka planda olasılıkları nasıl hesapladığı aşağıda açıklanmıştır. Burada kısaltma amacıyla $Gazete1=G1$, $Gazete2=G2$, $Gazete3=G3$ şeklinde sembolize edilecektir. Tablo 3.1a verisinden elde edilen her farklı gruptaki kişi sayısı Tablo 3.1b ile gösterilmiştir.

Tablo 3.1.a: Maaş-Gazete ilişkisi

Maaş	Gazete
Az	G1
Orta	G1
Yüksek	G1
Az	G2
Az	G3
.	.

Tablo 3.1.b: Her gruptaki kişi sayısı

Maaş	Gazete=G1	Gazete=G2	Gazete=G3
Az	60	10	30
Orta	20	30	20
Yüksek	4	6	20

Tablo 3.1b yardımıyla sözü edilen olasılıklar (3.8) ifadesi kullanılarak;

$$\begin{aligned}
 &P(\text{Gazete}=G_1 | \text{Maas} = A) \\
 &= \frac{P(G = G_1)P(M = A | G = G_1)}{P(G = G_1)P(M = A | G = G_1) + P(G = G_2)P(M = A | G = G_2) + P(G = G_3)P(M = A | G = G_3)} \\
 &= \frac{\frac{84}{200} \frac{60}{84}}{\frac{84}{200} \frac{60}{84} + \frac{46}{200} \frac{10}{46} + \frac{70}{200} \frac{30}{70}} = 0.6
 \end{aligned} \tag{3.11}$$

$$\begin{aligned}
 &P(\text{Gazete}=G_2 | \text{Maas} = A) \\
 &= \frac{P(G = G_2)P(M = A | G = G_2)}{P(G = G_1)P(M = A | G = G_1) + P(G = G_2)P(M = A | G = G_2) + P(G = G_3)P(M = A | G = G_3)} \\
 &= \frac{\frac{46}{200} \frac{10}{46}}{\frac{84}{200} \frac{60}{84} + \frac{46}{200} \frac{10}{46} + \frac{70}{200} \frac{30}{70}} = 0.1
 \end{aligned} \tag{3.12}$$

$$P(\text{Gazete}=G_3 | \text{Maas} = A) = 1 - (0.6 + 0.1) = 0.3 \tag{3.13}$$

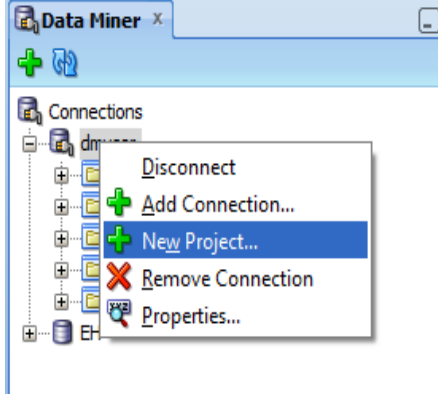
olarak hesaplanabilir. Burada bu sonuçlar çok daha kolay bir şekilde Tablo 3.1b’den de görülmektedir. Fakat hem hedef özelliğin ikiden fazla hem de kestirimci özellik sayısının birden fazla olduğu durumlarda tablodan okuma zorlaşacak ve (3.10) formülünün uygulanması gerekecektir. Benzer şekilde diğer olasılıklar da hesaplanarak;

$$\begin{aligned}
 &P(\text{Gazete}=G_1 | \text{Maas} = O) = 0.2857, & P(\text{Gazete}=G_2 | \text{Maas} = O) = 0.4286 \\
 &P(\text{Gazete}=G_2 | \text{Maas} = O) = 0.2857, & P(\text{Gazete}=G_1 | \text{Maas} = Y) = 0.1333 \\
 &P(\text{Gazete}=G_2 | \text{Maas} = Y) = 0.2, & P(\text{Gazete}=G_3 | \text{Maas} = Y) = 0.6667
 \end{aligned} \tag{3.14}$$

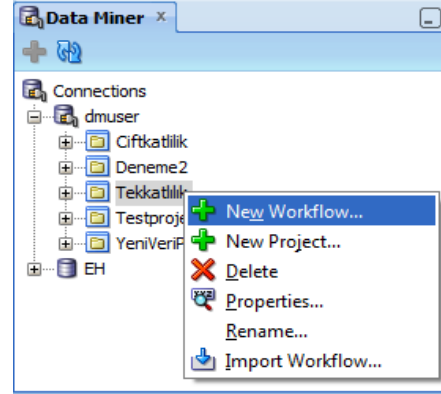
yazılabilir. Bu olasılıklar, gelir durumu bilinen birinin okuduğu gazete hakkında tahmin yapabilmeyi sağlayacaktır. Örneğin, maaş durumu az olan birinin okuduğu gazetenin Gazete1 olması beklenirken, maaş durumu yüksek olan birinin okuduğu gazetenin bu olasılıklara göre Gazete3 olması beklenecektir. Oracle Data Miner (ODM) Naive Bayes yöntemini kullanarak bu olasılıkları otomatik olarak hesaplayıp model oluşturmayı sağlar ve tahmin yapmayı kolaylaştırır.

Aynı işlem Oracle Data Miner kullanılarak yapılmıştır. İlk olarak “Data Miner” kullanıcısı olan “dmuser” içinde “New Project” ile ‘Proje_Hedef_Gazete’ adında

proje oluşturulmuş (Şekil 3.2), “New Workflow” sekmesiyle uygulamanın yer alacağı bir iş akış diyagramı eklenmiştir (Şekil 3.3) [6].

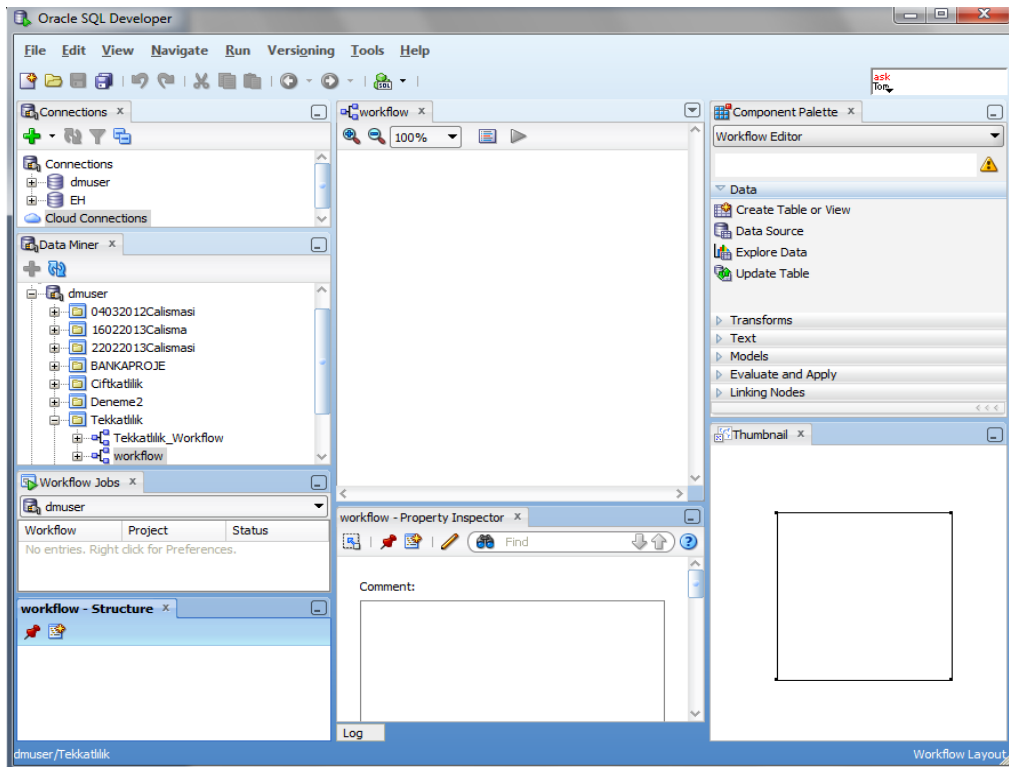


Şekil 3.2: Proje oluşturma



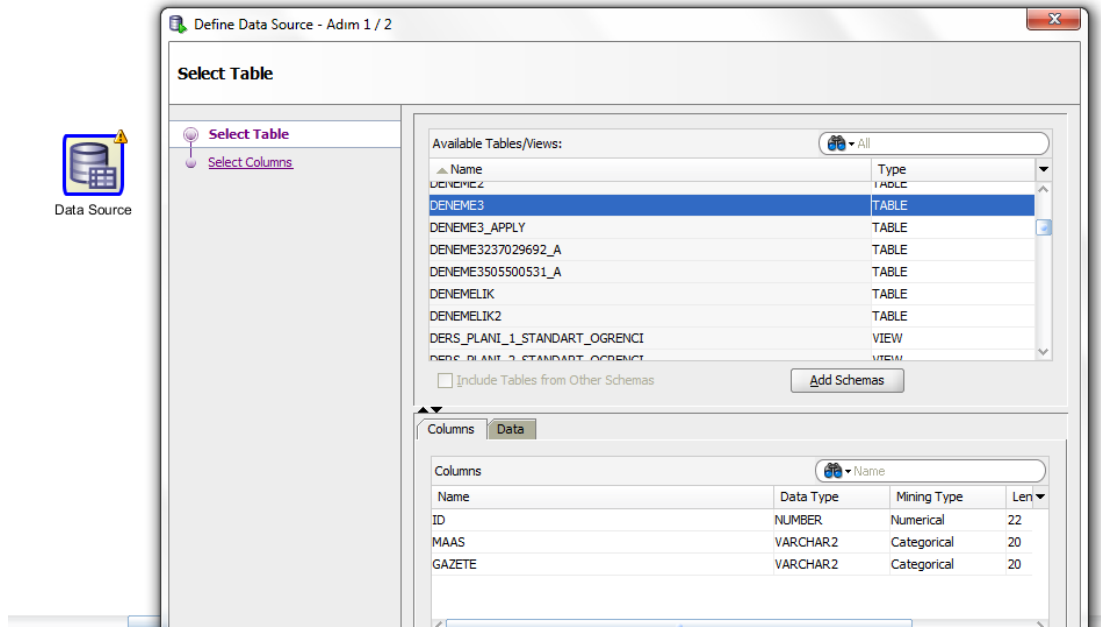
Şekil 3.3: “Workflow” oluşturma

Oluşturulan iş akış diyagramı Şekil 3.4’te görülmektedir. ODM arayüzünde “Component Palette” yer alır. Bu kısımda uygulama için kullanılabilecek tüm düğümler bulunmaktadır.



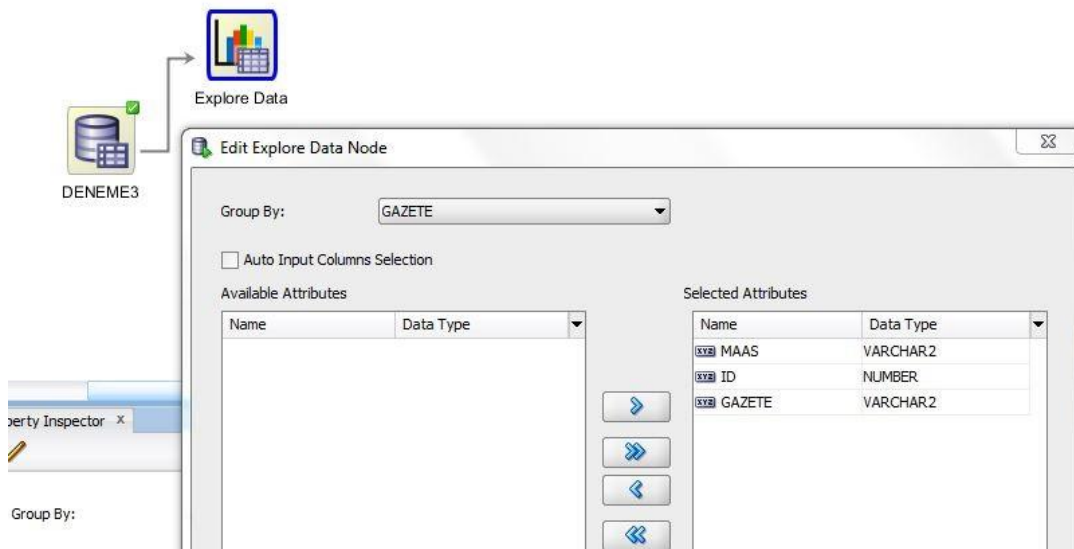
Şekil 3.4: ODM genel görünüm

İlk olarak “Data” sekmesi altından “Data Source” düğümü diyagram içine yerleştirilmiş ve olasılığı el ile hesaplanan, 100 kişinin maaş ve gazete bilgilerini içeren “DENEME3” adlı tablo gelen ekrandan seçilmiştir (Şekil 3.5).



Şekil 3.5: Tablo seçme

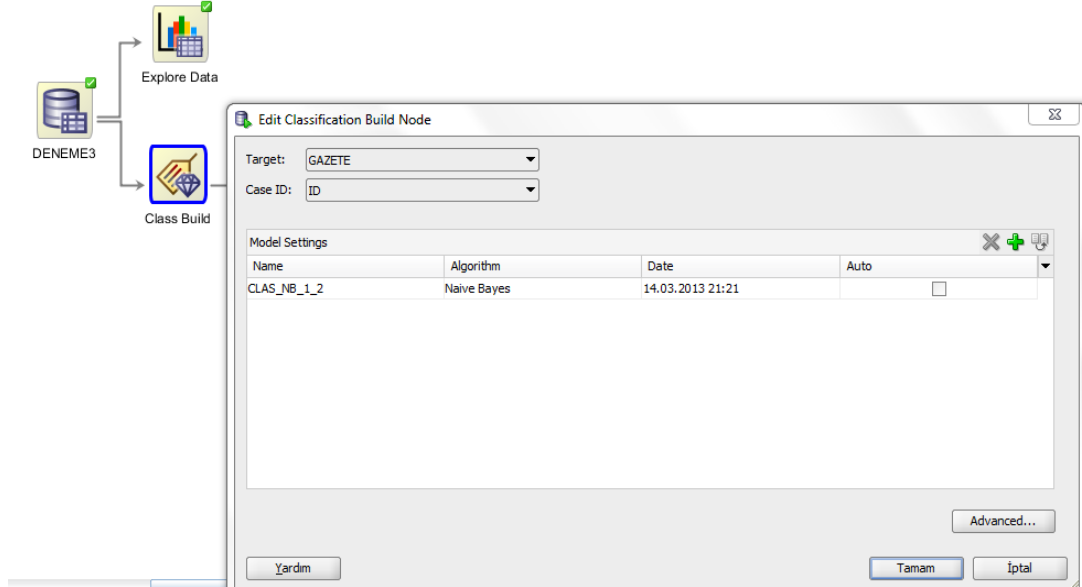
Veri seçildikten sonra, "Data" sekmesi altındaki "Explore Data" düğümü diyagramda yer alan veri ile bağlanmıştır. Burada kişilerin okudukları gazetelerin tahmin edilmesi amaçlandığından, verinin hedef sütuna göre gruplanması için Şekil 3.6'da görüldüğü gibi "Edit" seçeneğinden "Group By" kısmında 'GAZETE' seçilmiştir. "Explore Data" düğümü çalıştırılmıştır.



Şekil 3.6: "Explore Data" düzenleme

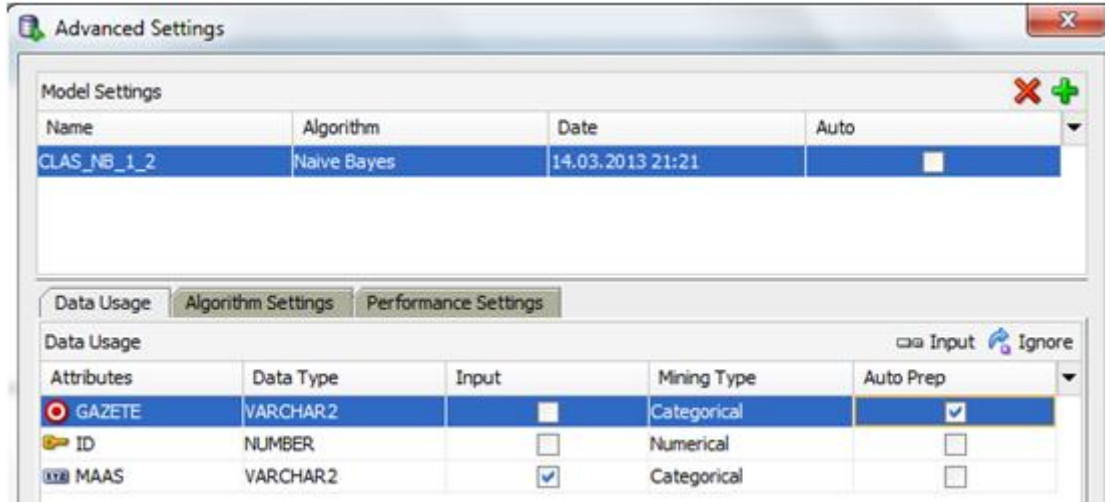
İşlem tamamlandığında düğüm üzerinde "View Data" seçilerek "Statics" bölümünde veri ile ilgili istatistikler incelenebilmektedir.

“Models” sekmesi altında yer alan “Classification” düğümü diyagramdaki veri kaynağı ile bağlanmış ve “Edit” sekmesinden yöntem için düzenleme yapılmıştır. Sınıflandırma işleminde sadece Naive Bayes algoritması kullanılmıştır. “Target” yani hedef sütun “GAZETE”, “CaseID” ise birincil anahtar olan “ID” seçilmiştir. Naive-Bayes algoritmasında ayarların otomatik olarak yapılmaması için “Auto” seçeneğini kaldırılmıştır (Şekil 3.7).



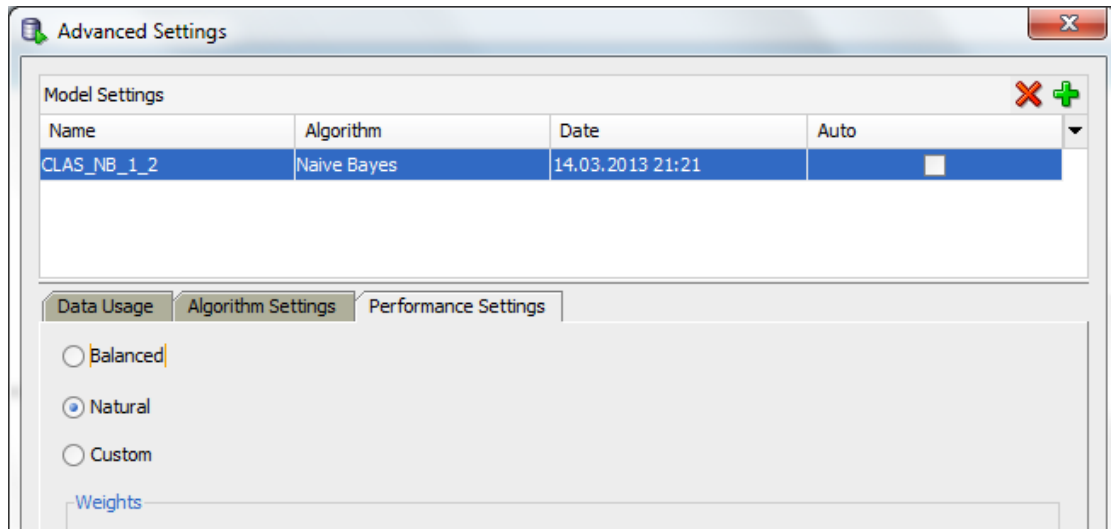
Şekil 3.7: “Classification” düğümü ayarları

Gelişmiş ayarları düzenlemek için için “Advanced Settings” bölümünde “Data Usage” sekmesinde “Input” ve “Ignore” butonları ile girdi değerleri düzenlenmiştir. Burada “ID” ve tahmin edilecek sütun olan “GAZETE”nin girdi olmamasına dikkat edilmiştir (Şekil 3.8). Girdi olarak alınan “MAAS” sütunu için “Auto Prep” (Auto Preparation) seçeneği kaldırılmıştır. Böylece veri üzerinde Naive Bayes algoritmasının herhangi bir hazırlık yapmasına izin verilmemiştir.



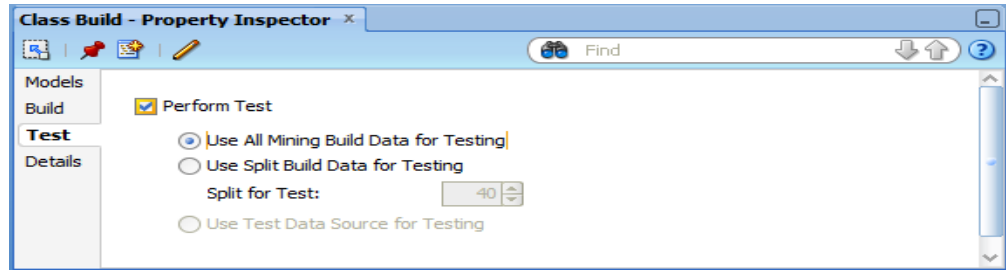
Şekil 3.8: “Data Usage” sekmesi

Ayrıca veri ağırlıklarının tabloda olduğu gibi yer alması için “Performance Settings” sekmesinde “Natural” seçeneği seçilerek ayarlar tamamlanmıştır (Şekil 3.9).



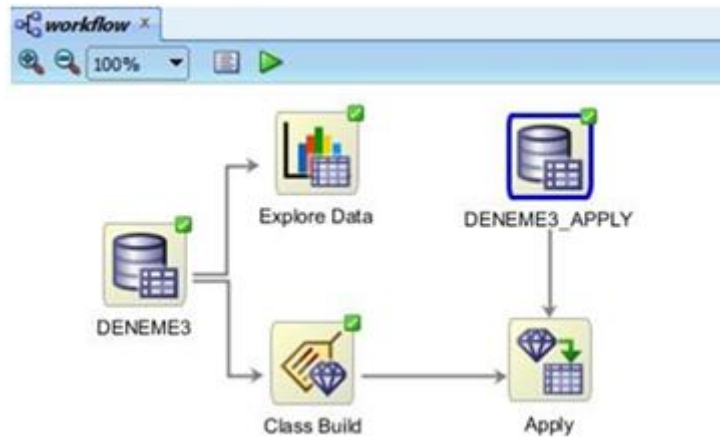
Şekil 3.9: “Performance Settings” sekmesi

"Classification" düğümü seçiliyken çıkan “Property Inspector” panosunda “Test” sekmesinde “Use All Mining Build Data for Testing” seçeneği seçilip (Şekil 3.10) "Classification" düğümü çalıştırılmıştır. Modelin testi için verinin tamamının kullanılmasının nedeni ODM tarafından yapılacak olasılık hesaplarının elle yapılan hesaplar ile aynı çıktığını görmektir.



Şekil 3.10: “Test” ayarları

Veri üzerinde tahmin modeli "Classification" ile oluşturulmuştur. Bu model kullanarak yapılan uygulama ile birlikte tahminler ve olasılık değerleri incelenebilmektedir. Bunun için “Evaluate and Apply” sekmesi altından “Apply” düğümü, iş akış diyagramına eklenmiştir. Oluşturulan model ve modelin oluşturulmasında kullanılan veri kaynağı, “Apply” düğümü ile Şekil 3.11’deki gibi bağlanarak “Apply” işleminde kullanılmıştır. Burada tahmin için aynı veri ele alınmıştır. Tahmin sonuçları olasılıklarla birlikte Şekil 3.9’da görülmüştür. Normal şartlarda ODM’de model oluşturulurken “Apply” düğümünde kullanılan veri “Classification” düğümünde kullanılanlardan farklı olmalıdır. Çünkü “Apply” aktivitesindeki amaç aynı sütunlara sahip farklı veriler içeren yeni veri için tahmin sütununun oluşturulmasıdır. Fakat burada ODM kullanmadan bulunan sonuçların doğruluğunun görülmesi amaçlandığından “Apply” düğümü aynı tabloya uygulanmıştır.



Şekil 3.11: Modelin uygulaması

Şekil 3.12’de görüldüğü gibi “CLAS_NB_1_2_PRED” isimli sütunda maaş durumlarına göre okunan gazete tahminleri yapılmıştır. Formül kullanarak bulunan sonuçlar ile “CLAS_NB_1_2_PROB” isimli sütundaki olasılıkları

karşılaştırıldığında iki olasılık değerlerinin aynı olduğu görülmüştür. Sonuç olarak el ile uzun işlemlerin yapılması yerine ODM kullanılarak daha büyük veriler için oluşturulan modeller ile tahmin yapılabilecektir.

workflow x

Apply x

Data

Columns | SQL

View: Actual Data

Sort...

Filter: Enter Where Clause

	ID	MAAS	GAZETE	CLAS_NB_1_2_PRED	CLAS_NB_1_2_PROB
	1	1 az	s	s	0,6
	2	2 az	s	s	0,6
	3	3 az	s	s	0,6
	4	4 az	s	s	0,6
	5	5 az	s	s	0,6
	6	6 az	s	s	0,6
	7	7 az	s	s	0,6
	8	8 az	s	s	0,6

154	154 orta	h	h	0,4286
155	155 orta	h	h	0,4286
156	156 orta	h	h	0,4286
157	157 orta	h	h	0,4286
158	158 orta	h	h	0,4286
159	159 orta	h	h	0,4286
160	160 orta	h	h	0,4286
161	161 orta	h	h	0,4286
162	162 orta	h	h	0,4286
163	163 orta	h	h	0,4286
164	164 orta	h	h	0,4286
165	165 orta	h	h	0,4286
166	166 orta	h	h	0,4286

181	181 yüksek	m	m	0,6667
182	182 yüksek	m	m	0,6667
183	183 yüksek	m	m	0,6667
184	184 yüksek	m	m	0,6667
185	185 yüksek	m	m	0,6667
186	186 yüksek	m	m	0,6667
187	187 yüksek	m	m	0,6667
188	188 yüksek	m	m	0,6667
189	189 yüksek	m	m	0,6667
190	190 yüksek	m	m	0,6667
191	191 yüksek	m	m	0,6667
192	192 yüksek	m	m	0,6667
193	193 yüksek	m	m	0,6667

Şekil 3.12 : Örnek 1'in ekran çıktıları

ODM, Naive Bayes yöntemiyle sınıflandırmaya mümkün olan her durum için olasılıkları hesaplayarak bir model oluşturup, bu modeli anlatıldığı gibi yeni

kayıtların durumunu tespit için kullanmaktadır. Modelin doğruluğunun test edilmesi amacıyla formüllerle yapılacak işlemlerde aynı veri ve hesaplanan olasılıklar kullanılarak yeni tahmin tablosu oluşturulabilir. Örneğin maaşı az olan kişinin Gazete1 gazetesi okuma olasılığı 0.60 olarak hesaplandığı için tahmin Gazete1 ve sonucun güvenilirliği 0.60'dır. Maaşı yüksek olan kişinin Gazete1 okuma olasılığı 0.1333 ve Gazete1 okuma olasılığı 0.20 olduğu için modelin tahmini Gazete3 ve sonucun güvenilirliği 0.6667 olacaktır. Bu şekilde işleme devam edilerek tüm tablo yeniden oluşturulur.

Tablo 3.2: Tablo 3.1a'nın yapılan hesaplamalarla elde edilen test sonuçları

Maaş	Gazete	Tahmin	Güvenilirlik
Az	Gazete1	Gazete1	0.60
Orta	Gazete2	Gazete2	0.4286
Yüksek	Gazete3	Gazete3	0.6667
Az	Gazete2	Gazete1	0.60
Orta	Gazete1	Gazete2	0.4286
	.	.	.

Modelin tüm güvenilirliği ise gerçek değerler ile tahmini değerlerin karşılaştırılması sonucu elde edilen aşağıdaki güvenilirlik matrisi ile verilebilir.

Tablo 3.3: Güvenilirlik Matrisi

	Gazete1	Gazete2	Gazete3
Gazete1	60	20	4
Gazete2	10	30	6
Gazete3	30	20	20

Tablo 3.3'te görülen güvenilirlik matrisinde satırlar gerçek değerleri, sütunlar ise tahmin sonuçlarını göstermektedir. Örneğin okuduğu gazete Gazete2 iken, modelin de Gazete2 olarak doğru tahmin ettiği kayıt sayısı 30, gerçekte okuduğu gazete Gazete2 iken modelin Gazete2 dışında bir gazete olarak yanlış tahmin ettiği kayıt sayısı 16'dır. Dolayısıyla matrisin köşegeni doğru kayıt sayısını, köşegen dışı ise yanlış kayıt sayısını göstermektedir. Buradan modelin doğruluğu

$$\frac{30 + 20 + 60}{30 + 6 + 10 + 20 + 20 + 30 + 20 + 4 + 60} = 0.65 \quad (3.15)$$

olarak elde edilir. Modelin güvenilirliği ODM kullanılarak da hesaplanabilir. "Classification" düğümü çalıştırıldıktan sonra "View Test Result" seçilerek Şekil 3.13'deki "Performance Matrix" kısmından modelin doğruluğu incelenebilir. "Overall Accuracy" modelin doğruluğunu göstermektedir. "Build", "Test" ve "Apply" işlemlerinde yukarıda anlatılan nedenlerden dolayı sadece örnekleme için aynı veri kullandığından el ile hesaplanan sonuçlarla ODM'nin hesapladığı sonuçların aynı çıktığı görülmektedir.

Target Value	Total Case Count	Correct Predictions %	Cost	Cost %
h	46	65,2174		
m	70	28,5714		
s	84	71,4286		

Şekil 3.13: Örnek 1'in "Performance Matrix" sonucu

Örnek 2) İki Boyut için: Yapılan bir anket sonucunda 200 denneğin gelir durumları "Az", "Orta" ve "Yüksek"; eğitim durumları ise "Lise", "Lisans" ve "Yüksek Lisans" olarak belirlenmiş ve bu özelliklerdeki deneklerin okudukları gazete ise "Gazete1", "Gazete2" ve "Gazete3" olarak belirlenip 3. bir sütunda verilmiş olsun (Tablo 3.4). Kayıtların dağılımı Tablo 3.5'te verilmiştir.

Tablo 3.4: Model verisi

Maaş	Eğitim	Gazete
Az	Lise	Gazete1
Orta	Lise	Gazete1
Yüksek	Lise	Gazete1
Az	Lise	Gazete2
Az	Lisans	Gazete2
.	.	.

Tablo 3.5 : Kayıt dağılımları

Maaş	Eğitim	Gazete	Kayıt Sayısı	Maaş	Eğitim	Gazete	Kayıt Sayısı
Az	Lise	Gazete1	10	Az	Lise	Gazete3	16
Az	Lisans	Gazete1	20	Az	Lisans	Gazete3	10
Az	Yüksek Lisans	Gazete1	30	Az	Yüksek Lisans	Gazete3	4
Orta	Lise	Gazete1	5	Orta	Lise	Gazete3	4
Orta	Lisans	Gazete1	7	Orta	Lisans	Gazete3	7
Orta	Yüksek Lisans	Gazete1	8	Orta	Yüksek Lisans	Gazete3	9
Yüksek	Lise	Gazete1	2	Yüksek	Lise	Gazete3	5
Yüksek	Lisans	Gazete1	1	Yüksek	Lisans	Gazete3	8
Yüksek	Yüksek Lisans	Gazete1	1	Yüksek	Yüksek Lisans	Gazete3	7
Az	Lise	Gazete2	2	Yüksek	Lise	Gazete2	1
Az	Lisans	Gazete2	5	Yüksek	Lisans	Gazete2	2
Az	Yüksek Lisans	Gazete2	3	Yüksek	Yüksek Lisans	Gazete2	3
Orta	Lise	Gazete2	5				
Orta	Lisans	Gazete2	10				
Orta	Yüksek Lisans	Gazete2	15				

Tablo 3.5 yardımıyla deneklerin gelir verileri de kullanılarak hangi gazeteyi okudukları olasılıkları (3.10) ifadesiyle belirlenmeye çalışılmıştır. Burada iki belirleyici özellik olduğundan (3.10) ifadesi bu tabloya uygun formda yazılmalıdır. Burada gazete olarak G1= Gazete1, G2=Gazete2, G3= Gazete3; maaş olarak A=Az, O=Orta, Y=Yüksek; eğitim olarak Lise=L, Lisans=Li, YL=Yüksek Lisans'ı ifade etmektedir.

(3.10) ifadesi örneğe uygulandığında,

$$\begin{aligned}
 & P(G=G_1|M=A, E=Li) \\
 &= \frac{P(G=G_1)P(M=A|G=G_1)P(E=Li|G=G_1)}{P(G=G_1)P(M=A|G=G_1) + P(G=G_2)P(M=A|G=G_2) + P(G=G_3)P(M=A|G=G_3)} \\
 &= \frac{\frac{84}{200} \frac{60}{84} \frac{17}{84}}{\frac{84}{200} \frac{60}{84} \frac{17}{84} + \frac{46}{200} \frac{10}{46} \frac{8}{46} + \frac{70}{200} \frac{30}{70} \frac{25}{70}} = 0.4937
 \end{aligned} \tag{3.16}$$

$$\begin{aligned}
 & P(G=G_2|M=A, E=Li) \\
 &= \frac{P(G=G_2)P(M=A|G=G_2)P(E=Li|G=G_2)}{P(G=G_1)P(M=A|G=G_1) + P(G=G_2)P(M=A|G=G_2) + P(G=G_3)P(M=A|G=G_3)} \\
 &= \frac{\frac{46}{200} \frac{10}{46} \frac{8}{46}}{\frac{84}{200} \frac{60}{84} \frac{17}{84} + \frac{46}{200} \frac{10}{46} \frac{8}{46} + \frac{70}{200} \frac{30}{70} \frac{25}{70}} = 0.0707
 \end{aligned} \tag{3.17}$$

$$P(G=G_3|M=A, E=Li) = 1 - (0.4937 + 0.0707) = 0.4356 \tag{3.18}$$

sonucu elde edilir. Bu sonuca göre geliri düşük ve üniversite mezunu olan kişiler %43.56 olasılıkla Gazete3 gazetesi okuyan gruptan olacaktır. Benzer şekilde diğer bazı olasılıklar da hesaplanırsa;

$$\begin{aligned}
 & P(G=G_1|M=A, E=Li) = 0.4937, & P(G=G_1|M=A, E=L) &= 0.5812, \\
 & P(G=G_1|M=A, E=YL) = 0.6795, & P(G=G_3|M=O, E=Li) &= 0.4353, \\
 & P(G=G_2|M=O, E=L) = 0.4453, & P(G=G_2|M=O, E=YL) &= 0.4773, \\
 & P(G=G_3|M=Y, E=Li) = 0.794, & P(G=G_3|M=Y, E=L) &= 0.668, \\
 & P(G=G_3|M=Y, E=YL) = 0.5542
 \end{aligned} \tag{3.19}$$

Olasılıkların uzun işlemlerle hesaplanıp, tahminde bulunulması yerine, aynı işlem ODM üzerinde Örnek 1’de belirtilen adımlarla yapılmıştır. Örnek 2’nin “Apply” işleminin sonucu tahminler ve olasılıklar ile birlikte Şekil 3.14’te gösterilmiştir.

Gazete Belirleme x Apply x					
Data	Columns	SQL			
View: Actual Data	Sort...	Filter: Enter Where Clause			
ID	MAAS	EGITIM	GAZETE	CLAS_NB_1_5_PRED	CLAS_NB_1_5_PROB
1	1 az	lise s	s		0,4937
2	2 az	lise s	s		0,4937
3	3 az	lise s	s		0,4937
4	4 az	lise s	s		0,4937
5	5 az	lise s	s		0,4937
6	6 az	lise s	s		0,4937
7	7 az	lise s	s		0,4937
8	8 az	lise s	s		0,4937
9	9 az	lise s	s		0,4937
10	10 az	lise s	s		0,4937
45	45 az	lisans s	s		0,5812
46	46 az	lisans s	s		0,5812
47	47 az	lisans s	s		0,5812
48	48 az	lisans s	s		0,5812
49	49 az	lisans m	s		0,5812
50	50 az	lisans m	s		0,5812
51	51 az	lisans m	s		0,5812
52	52 az	lisans m	s		0,5812
53	53 az	lisans m	s		0,5812
54	54 az	lisans m	s		0,5812
55	55 az	lisans m	s		0,5812
56	56 az	lisans m	s		0,5812
57	57 az	lisans m	s		0,5812
58	58 az	lisans m	s		0,5812
59	59 az	lisans h	s		0,5812
60	60 az	lisans h	s		0,5812
61	61 az	lisans h	s		0,5812
177	177 yüksek	lise m	m		0,794
178	178 yüksek	lise h	m		0,794
179	179 yüksek	lisans s	m		0,668
180	180 yüksek	lisans m	m		0,668
181	181 yüksek	lisans m	m		0,668
182	182 yüksek	lisans m	m		0,668
183	183 yüksek	lisans m	m		0,668
184	184 yüksek	lisans m	m		0,668
185	185 yüksek	lisans m	m		0,668
186	186 yüksek	lisans m	m		0,668
187	187 yüksek	lisans m	m		0,668
188	188 yüksek	lisans h	m		0,668
189	189 yüksek	lisans h	m		0,668
190	190 yüksek	yl s	m		0,5542
191	191 yüksek	yl m	m		0,5542
192	192 yüksek	yl m	m		0,5542

Şekil 3.14: Örnek 2'nin ekran çıktıları

Ayrıca Örnek 2’deki modelin doğruluğu Şekil 3.15’te gösterilmiştir.

Performance Performance Matrix Lift Profit					
Average Accuracy: 53,354					
Overall Accuracy: 54,5					
Target Value	Total Case Count	Correct Predictions %	Cost	Cost %	
h	46	54,3478			
m	70	34,2857			
s	84	71,4286			

Şekil 3.15: Örnek 2’nin “Performance Matrix” sonucu

3.2 Birliktelik Kuralları (Association Rules - AR)

Olayların birlikte gerçekleşme durumlarını çözümleyen veri madenciliği yöntemlerine birliktelik kuralları denir. Bu çözümlerler çoğunlukla müşterilerin satın alma eğilimlerini belirlemek amacıyla yapılmaktadır. Satın alma eğilimleri müşterilerin bir arada satın aldığı tüm ürünleri ele alarak ortaya konur, bu uygulamalara “pazar sepet çözümü” denmektedir [4].

3.2.1 Destek ve güven ölçütleri

Pazar-sepet çözümlerinde ilişkiler için destek ve güven ölçütlerinden yararlanılır. Destek sayısı adı verilen değer ise bu ölçütlerin hesaplanmasında kullanılır. Bir ilişkinin tüm alışverişler içinde hangi oranda tekrarlandığını belirlemek için kural destek ölçütü kullanılır. X ürün grubunu alan müşterilerin Y ürün grubunu da alma olasılığını kural güven ölçütü ortaya koyar. Bu durum $X \rightarrow Y$ biçiminde gösterilir ve şu şekilde ifade edilebilir.

$$destek(X \rightarrow Y) = \frac{sayi(X, Y)}{N} \quad (3.20)$$

Burada X ve Y ürün gruplarını birlikte içeren alışveriş sayısını $sayi(X, Y)$, tüm alışveriş sayısını ise N göstermektedir.

$$guven(X \rightarrow Y) = \frac{sayi(X, Y)}{sayi(X)} \quad (3.21)$$

$sayi(X)$ ise X ürün grubu alışveriş sayısını ifade etmektedir.

Bu ölçütlerin yanı sıra, bu değerleri karşılaştırmak üzere eşik değere gereksinim vardır. Hesaplanan destek ve güven ölçütleri eşik değerlerden büyük olmalıdır. Bu ölçütler ne kadar büyük ise birliktelik kuralları da o derece güçlüdür.

3.2.2 Örnek

Bir mağazada 100 müşterinin bir defada yaptığı alışveriş bilgilerinden yararlanarak birliktelik kurallarının şu şekilde elde edildiğini varsayalım.

$$guven(seker, cay \rightarrow ekmek) \quad (3.22)$$

Burada $X = \{seker, cay\}$ ürünlerini birlikte alan müşterilerin $Y = \{ekmek\}$ ürününü de satın alma olasılığını ifade etmektedir.

$\{seker, cay, ekmek\}$ ile ilgili destek sayısı, yani bu üç ürünü birlikte satın alma sayısı 30 ve müşteri sayısı 100 ise

$$destek(seker, cay \rightarrow ekmek) = \frac{sayi(seker, cay, ekmek)}{musteri sayisi} = \frac{30}{100} = \%30 \quad (3.23)$$

ile hesaplanır.

Bu kez $\{seker, cay\}$ ile ilgili destek sayısının, yani bu iki ürünü birlikte satın alanların sayısının 40 olduğunu varsayalım. O halde kural güven ölçütü şu şekilde ifade edilir:

$$guven(seker, cay \rightarrow ekmek) = \frac{sayi(seker, cay, ekmek)}{sayi(seker, cay)} = \frac{30}{40} = \%75 \quad (3.24)$$

3.2.3 Apriori algoritması

Birliktelik kurallarının üretilmesi için kullanılan yöntemlerden en yaygını "Apriori algoritması" dır. Bu algoritma aşağıda belirtilen aşamalara sahiptir:

1. Birliktelik çözümlemesinin yapılabilmesi için öncelikle destek ve güven ölçütlerini karşılaştırmak üzere eşik değerler belirlenir. Uygulamadan elde edilen sonuçların bu eşik değerlere eşit ya da büyük olması gerekir.
2. Veri tabanı taranarak çözümlemeye dahil edilecek her bir ürün için destek sayıları hesaplanır. Bu destek sayıları eşik destek sayısı ile karşılaştırılır. Eşik destek

sayısından küçük değerlere sahip satırlar çözümlemeyen çıkarılır ve koşula uygun kayıtlar göz önüne alınır.

3. Yukarıdaki adımda seçilen ürünler bu kez ikiye bölünebilir gruplandırılarak, bu grupların destek sayıları elde edilir. Bu sayılar eşik destek sayıları ile karşılaştırılır ve eşik değerden küçük değerlere sahip satırlar çözümlemeyen çıkarılır.

4. Bu kez üçerli, dörderli vb. gruplandırmalar yapılarak bu grupların destek sayıları elde edilir ve eşik değer ile karşılaştırılır. Eşik değerlere uygun olduğu sürece işlemlere devam edilir.

5. Ürün grubu belirlendikten sonra kural destek ölçütüne bakılarak birliktelik kuralları türetilir ve bu kuralların her birisiyle ilgili olarak güven ölçütleri hesaplanır.

3.3 Özelliklerin Önem Sıralaması (Attribute Importance - AI)

ODM kullanıcılarına, her özelliğin yani her tahmin edici kolonun sonuç kolonu üzerinde etkisini “En Küçük Tanımlama Uzunluğu (Minimum Description Length-MDL)” adı verilen bir yöntemle ölçerek elde edilen ölçüm değerlerine göre tahmin edici kolonları yani özellikleri önem sırasına göre sıralayan bir uygulama sunmaktadır [5]. Özellikle tahmin edici sütun sayısının çok fazla olduğu durumlarda, birçok kolonun hedef kolon üzerinde etkisi olmayabilir hatta bu kolonlar modelin doğruluğuna negatif etki edebilir. Model oluşturma safhasında negatif etki yapan kolonların modelden çıkarılmasının doğruluğu artıracağı açıktır. Bu amaçla her özelliğin hedef sütunun belirlenmesine katkısı MDL prensibi ile belirlenebilir; önemini hesaplamak istediğimiz sütunun i ile belirtilen her farklı değerinin n_i kayıt içerdiğini ve buna karşılık hedef kolonun m farklı sınıf içerdiğini, ve bunların olma olasılıklarının p_{ij} ile verildiğini kabul edelim. Bu durumu açıklamak için yukarıdaki örnekten faydalanalım; Tahmin edici sütun “Maaş” sütunu ve hedef sütun okunan gazete yani “Gazete” sütunu olsun, “Maaş” sütununda toplam 200 kayıt bulunduğunu varsayalım. “Maaş” sütununda “Az”, “Orta” ve “Yüksek” olmak üzere üç farklı sınıf bulunmaktadır ve bu sınıfların her birinin sırasıyla 100, 70, 30 kayıt içerdiğini kabul edelim. Burada sırasıyla $i=1$ “Az”, $i=2$ “Orta”, $i=3$ “Yüksek” sınıflarına karşı gelmekte ve dolayısıyla $n_1=100$, $n_2=70$, $n_3=30$ olmaktadır. Hedef sütunda toplam m farklı sınıf ise örnekte “Gazete1”, “Gazete2” ve “Gazete3” olmak üzere

$m = 3$ farklı sınıfa karşı gelmektedir. i . farklı değere karşı gelen hedef kolondaki j . farklı sınıfın olma olasılığı p_{ij} ile verilmektedir. Eğer i . farklı değer için n_i örnekten her birine hedef kolonda n_i tane farklı değer karşı gelseydi ki, (örneğimizde bu mümkün değildir çünkü bu durumda $m \geq n_i$ olması gerekirdi) en fazla n_i p_{ij} farklı durum görülebilirdi. Kombinasyon hesabından toplam n objenin en fazla m farklı gruba $Kombinasyon(n+m-1, m-1)$ şekilde ayrılabilceği bilinir. Örneğin $n = 4$ ve $m = 2$ ise, toplam 4 nesne ile en fazla iki farklı grup oluşturabilecek dağılım sayısı $Kombinasyon(4+2-1, 2-1) = Kombinasyon(5,1) = 5$ şeklindedir ve bu gruplar “AAAA”, “AAAB”, “AABB”, “ABBB” ve “BBBB” şeklindedir. Ayrıca "Shannon gürültüsüz kodlama" teoreminden, bir x ifadesinin, en kompakt şekilde şifrenmesi için gereken bit sayısı $\log_2 x$ ile verilir. Dolayısıyla “Model penalty formüle” olarak isimlendirilen bir tahmin edici kolonun boyutu

$$\sum_i \log_2 [Comb(n+m-1, m-1)] \quad (3.25)$$

ile verilir. Verilen olasılıklar ile birlikte bir tahmin edici kolonun toplam uzunluğu

$$MDL = \sum_i \log_2 [Comb(n+m-1, m-1)] - \sum_i \log_2 P_{ij} \quad (3.26)$$

olarak verilir. (3.10) yardımıyla her özellik için bir MDL değeri elde edilir ki, bu değerlere göre tahmin edici kolonların önemi belirlenir, örnek Şekil 3.15’te görülmektedir.

Attribute Importance			
Data Columns SQL			
Target: AFFINITY_CARD			
Attribute Ranking			
Name	Type	Rank	Importance
HOUSEHOLD_SIZE	VARCHAR2	1	0,1589
CUST_MARITAL_STATUS	VARCHAR2	2	0,1582
YRS_RESIDENCE	NUMBER	3	0,0941
EDUCATION	VARCHAR2	4	0,0863
AGE	NUMBER	5	0,0849
OCCUPATION	VARCHAR2	6	0,0752
Y_BOX_GAMES	NUMBER	7	0,063
HOME_THEATER_PACKAGE	NUMBER	8	0,0565
CUST_GENDER	CHAR	9	0,0353
BOOKKEEPING_APPLICATION	NUMBER	10	0,0192
BULK_PACK_DISKETTES	NUMBER	11	0
COUNTRY_NAME	VARCHAR2	11	0
CUST_ID	NUMBER	11	0

Şekil 3.16 : Oracle Data Miner ile belirlenmiş AI değerleri

4. İSTATİSTİKLER

Bu bölümde İstanbul Teknik Üniversitesi Öğrenci İşleri Daire Başkanlığı'ndan elde edilen Matematik Mühendisliği öğrenci verileri yardımıyla çıkarılan istatistiklere yer verilmiştir. Burada yer verilen tüm istatistikler, mezun standart ÖSS/YÖS kabul tipindeki öğrencilerin bilgilerini baz alarak hesaplanmıştır.

Tablo 4.1: Mezuniyet Dönem Sayısı – Öğrenci Sayısı İstatistikleri

Mezuniyet Dönem Sayısı	Öğrenci Sayısı	Mezuniyet Dönem Sayısı	Öğrenci Sayısı
6	1	14	9
7	5	15	1
8	39	16	0
9	69	17	0
10	54	18	2
11	41	19	0
12	23	20	1
13	14		

Tablo 4.1'den görülür ki en fazla bitirme dönemi sayısı 20 ve en az bitirme dönemi sayısı 6'dır. Standart ÖSS/YÖS kabul tipindeki öğrencilerden sadece 1 kişi 6 dönemde öğrenimini bitirmiştir. Bu öğrenci 330 ID'li öğrencidir ve öğrencinin transfer bilgilerinde 41 kredilik ders olduğu görülmüştür. Bu öğrencinin daha önce İstanbul Teknik Üniversitesi'nde başka bir bölümde okuduğu ve o bölümde başardığı derslerden 41 kredilik kısmını Matematik Mühendisliği bölümü dersleri yerine saydırdığı anlaşılmaktadır. 7 dönemde bitiren öğrencilerden 73 ID'li öğrencinin bilgilerine bakıldığında, sadece güz ve bahar dönemlerinde ders alarak öğrenimini bitirdiği görülür. Ayrıca bu öğrencinin sadece bir dönemde ortalaması 3.00'ın altında kalmıştır. 149, 227 ve 232 ID'li öğrenci sadece güz ve bahar dönemlerinde ders almıştır ve ortalaması her dönem 3.00 üzerindedir. 241 ID'li öğrencinin, düzensiz

öğrenci olduğu görülür. İki dönemde ortalaması 2.90 üzeri ve diğer dönemlerde ortalaması 3.00 üzeridir.

Tablo 4.2: Ders tekrar sayısı istatistikleri

DERS	ORTALAMA TEKRAR SAYISI	MAKSİMUM TEKRAR SAYISI	DERS	ORTALAMA TEKRAR SAYISI	MAKSİMUM TEKRAR SAYISI
MAT213	1,15	4	ATA101	1,065455	2
MAT214	1,294545	4	ATA102	1,08	2
MAT232	1,261818	5	MAT492	1,003636	2
MAT322E	1,276364	4	MAT111	1,214545	4
MAT221	1,487273	4	MAT112	1,225455	4
MAT242	1,570909	5	MAT141	1,334545	4
MUH211	1,061818	4	MAT142	1,236364	4
MUH212	1,210909	3	MAT175	1,196364	3
DNK201E	1,258182	5	FİZ101E	1,189091	6
ING201	1,010909	2	FİZ101EL	1,010909	2
TUR101	1,020901	2	FİZ106	1,276364	4
TUR102	1,061818	2	FİZ106L	1,007273	2
MAT331	1,338182	4	KİM101E	1,112727	3
MAT341E	1,16	3	KİM101EL	1,007273	2
MAT332	1,167273	3	BİL101E	1,727273	7
MAT342E	1,156364	3	BİL104E	1,189091	3
MUH311	1,338182	6	ING101	1,036364	2
MUH312	1,32	5	ING102	1,054545	3
MUH321	1,069091	2			

Tablo 4.2’den bölüm derslerinden sırasıyla MAT492, MUH211 ve MUH321’in en az tekrar edilen üç ders olduğu ve sırasıyla MAT242, MAT211 ve MUH311’in en fazla tekrar edilen üç ders olduğu görülmektedir.

Tablo 4.3’ten Tablo 4.34’e kadar BİL101E dersi de dahil olmak üzere 2 kredi üzerindeki tüm dersler için dersi alan sayısı, dersin ortalaması bilgileri verilmiştir (Ek-14), bu tablolar için sütun açıklamaları şöyledir:

Sadece 1 Kez: “Dersi sadece 1 kez alanların sayısı”

1.Ort: “Dersin dersi sadece 1 kez alanlara göre ortalaması”

İlkKez: “Dersi ilk kez alanların sayısı (zorunlu dersler için öğrenci sayısına eşittir)”

İlkKezOrt: ”Dersin dersi ilk kez alanlara göre ortalaması”

Sadece2Kez: “Dersi sadece 2 kez alanların sayısı”

2.Ort: “Dersi sadece 2 kez alanların 2. kez aldıkları notların ortalaması”

EnAz2Kez: “Dersi en az 2 kez alanların sayısı”

İkinciKezOrt: “Dersi en az 2 kez alanların 2. kez aldıkları notların ortalaması”

Sadece3Kez: “Dersi sadece 3 kez alanların sayısı”

3.Ort: “dersi sadece 3 kez alanların 3. kez aldıkları notların ortalaması”

EnAz3Kez: “Dersi en az 3 kez alanların sayısı”

UcuncuKezOrt: “Dersi en az 3 kez alanların 3. kez aldıkları notların ortalaması”

Tablo 4.3: MAT111 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Ucuncu Ort	En Az 3 Kez	Ucuncu Kez Ort
223	2.8116	275	2.4909	47	2.946	52	2.663	3	1.8333	5	1.1

Tablo 4.4: MAT112 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Ucuncu Ort	En Az 3 Kez	Ucuncu Kez Ort
223	2.7376	275	2.3654	43	2.5814	52	2.173	8	2.25	9	2.

Tablo 4.5: MAT141 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Ucuncu Ort	En Az 3 Kez	Ucuncu Kez Ort
191	2.4869	275	1.9745	78	2.673	84	2.482	4	1.75	6	1.167

Tablo 4.6: MAT142 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Ucuncu Ort	En Az 3 Kez	Ucuncu Kez Ort
217	2.3525	274	2.0219	51	2.803	57	2.561	4	2.875	6	2.333

Tablo 4.7: MAT175 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Ucuncu Ort	En Az 3 Kez	Ucuncu Kez Ort
222	2.2545	275	1.9218	52	2.490/	53	2.443	1	3.	1	3.

Tablo 4.8: FİZ101E için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
230	2.4913	275	2.1745	41	2.463	45	2.277	3	2.6666	4	2.

Tablo 4.9: FİZ106 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
211	2.3199	275	1.8618 22	54	2.888 8	64	2.437 5	8	2.437	10	1.95

Tablo 4.10: KİM101E için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
245	2.6040 8	275	2.4454 5	29	2.517 24	30	2.433 33	1	2.	1	2.

Tablo 4.11: BİL101E için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
158	1.7658	275	0.2945 5	64	- 0.141	117	- 0.965 8	34	1.0882 4	53	0.207 55

Tablo 4.12: BİL104E için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
225	2.8311 1	275	2.4072 7	48	2.833 33	50	2.72	2	2.75	2	2.75

Tablo 4.13: İNG101 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
265	3.2962	275	3.2182	10	3.5	10	3.5	0	İndeterminate	0	İndeterminate

Tablo 4.14: İNG102 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
262	3.4045 8	275	3.2709	11	3.227 27	13	2.730 77	2	3.	2	3.

Tablo 4.15: MAT213 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
239	2.7008	275	2.4127	32	2.937	36	2.68	3	1.8333	4	1.375

Tablo 4.16: MAT214 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
206	2.1796	275	1.6981	59	2.483	69	2.123	8	1.8125	10	1.45

Tablo 4.17: MAT221 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
162	2.2098	275	1.36	97	2.469	113	2.119	11	2.1363	16	1.468

Tablo 4.18: MAT232 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
212	2.4174	275	1.9581	57	2.535	63	2.293	4	2.125	6	1.416

Tablo 4.19: MAT242 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
151	2.2086	275	1.3036	100	2.51	124	2.072	17	2.6764	24	1.895

Tablo 4.20: MUH211 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
260	2.564	275	2.44	14	2.392	15	2.233	0		1	0.

Tablo 4.21: MUH212 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
224	2.3258	275	1.9291	44	2.295	51	1.980	7	2.0714	7	2.071

Tablo 4.22: DNK201E için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
212	2.6792	275	2.1236	58	2.905	63	2.674	3	2.6666	5	1.6

Tablo 4.23: ING201 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
272	3.798	275	3.756	3	3.67	3	3.67	0		0	

Tablo 4.24: MAT322E için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
214	2.1448	275	1.7036	48	2.156	61	1.696	11	2.2727	13	1.923

Tablo 4.25: MAT331 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
197	2.2208	275	1.6527	67	2.194	78	1.923	7	1.7143	11	1.090

Tablo 4.26: MAT332 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
234	2.4059	275	2.0672	36	2.375	41	2.085	5	2.1	5	2.1

Tablo 4.27: MAT341E için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
233	2.4742	275	2.1418	40	2.3	42	2.190	2	1.75	2	1.75

Tablo 4.28: MAT342E için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
234	2.666	275	2.374	39	2.78	41	2.64	2	1.5	2	1.5

Tablo 4.29: MAT355 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
220	2.3364	259	2.0270	29	2.310	39	1.717	7	2.	10	1.95

Tablo 4.30: MUH311 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
211	2.1659	275	1.6855	43	2.267	64	1.523	15	1.5	21	1.119

Tablo 4.31: MUH312 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
204	2.0269	275	1.5218	61	2.221	71	1.908	5	2.	10	1.1

Tablo 4.32: MUH321 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
254	2.8799	275	2.7473	19	2.868	21	2.595	2	3.5	2	3.5

Tablo 4.33: EKO201 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
262	2.7061	274	2.6295	10	2.65	12	2.208	2	3.25	2	3.25

Tablo 4.34: MAT492 için dersi alan sayısı, dersin ortalaması

Sadece 1 Kez	Birinci Ort	İlk Kez	İlk Kez Ort	Sadece 2 Kez	İkinci Ort	En Az 2 Kez	İkinci Kez Ort	Sadece 3 Kez	Üçüncü Ort	En Az 3 Kez	Üçüncü Kez Ort
274	3.9762	275	3.9618	1	4.	1	4.	0		0	

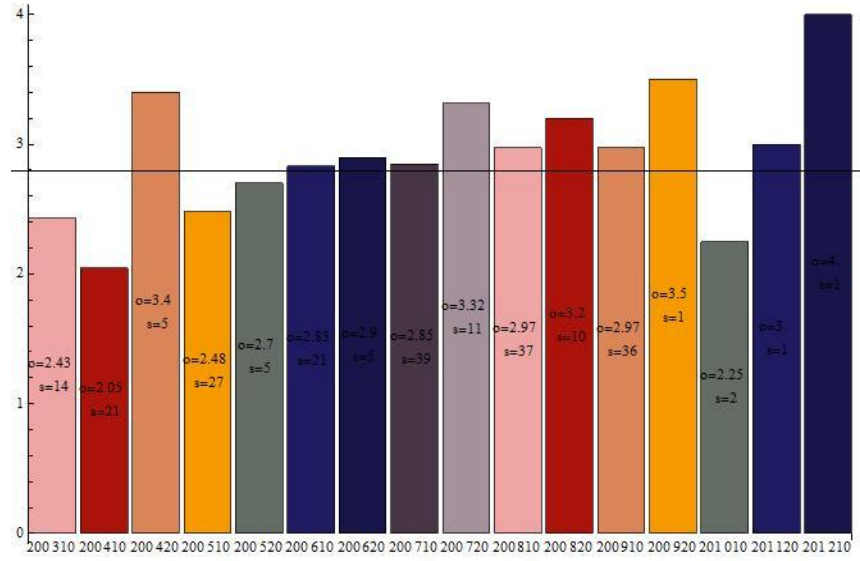
Tablo 4.35: Dersi ilk kez alanlara göre bölüm derslerinin azalan sırada ortalamaları

DERS	ORTALAMA
MAT242	1,3
MAT221	1,36
MUH312	1,52
MAT331	1,65
MUH311	1,69
MAT214	1,7
MAT322E	1,7
MAT175	1,92
MUH212	1,93
MAT232	1,96
MAT141	1,98
MAT142	2,02
MAT332	2,07
MAT341E	2,14
MAT112	2,37
MAT342E	2,37
MAT213	2,41
MUH211	2,44
MAT111	2,49
MUH321	2,75
MAT492	3,96

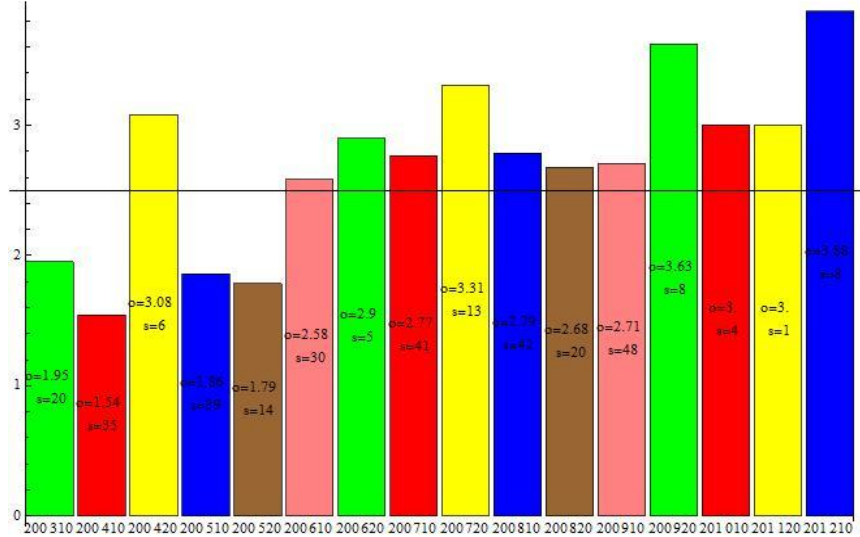
Tablo 4.35'ten görüldüğü üzere en düşük ilk üç ortalamaya sahip bölüm dersleri sırasıyla MAT242, MAT221, MUH312 olduğundan bu derslere en zor bölüm dersleridir denilebilir. En yüksek ilk üç ortalamaya sahip bölüm dersleri ise sırasıyla

MAT492, MUH321, MAT111 olduğundan bu derslere en kolay bölüm dersleridir denilebilir.

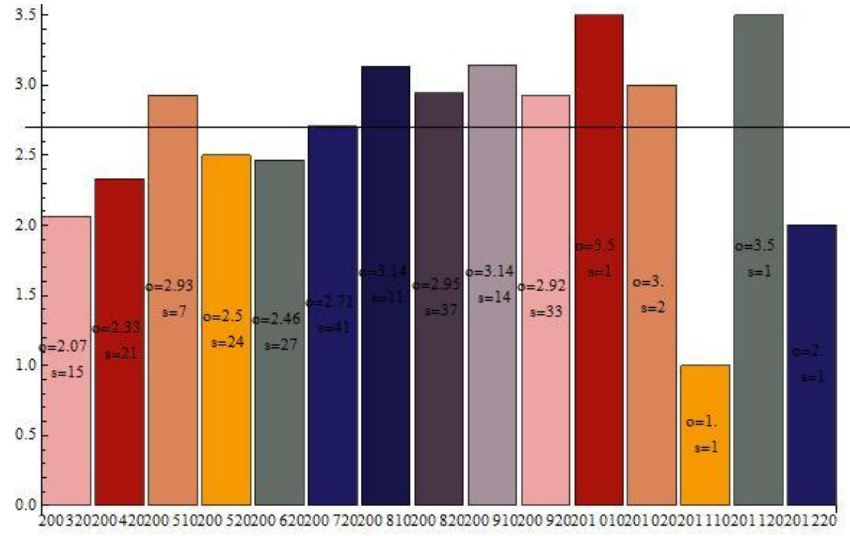
Şekil 4.1’den Şekil 4.64’e kadar verilen sütun grafiklerinde BIL101E dahil 2 kredi ve üzerindeki tüm derslerin dersi ilk kez alan öğrencilere ve dersi alan tüm öğrencilere göre dönem dönem ortalamaları verilmiştir, grafiklerde düz çizgiyle gösterilen eşik değeri tüm dönemlerin ortalamasına göre belirlenmiştir. Her bir sütun bir dönemi göstermektedir, sütunların üzerlerinde “o” ile dersin dönemdeki ortalaması, “s” ile dönemde dersi alan öğrenci sayısı verilmiştir (Ek-15).



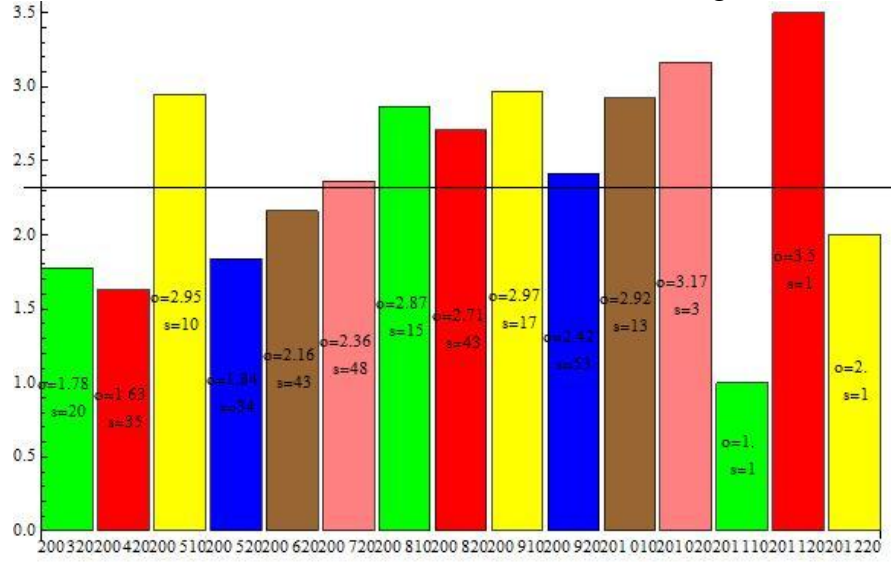
Şekil 4.1: MAT111 dersini ilk kez alanların dönemlere göre ortalaması



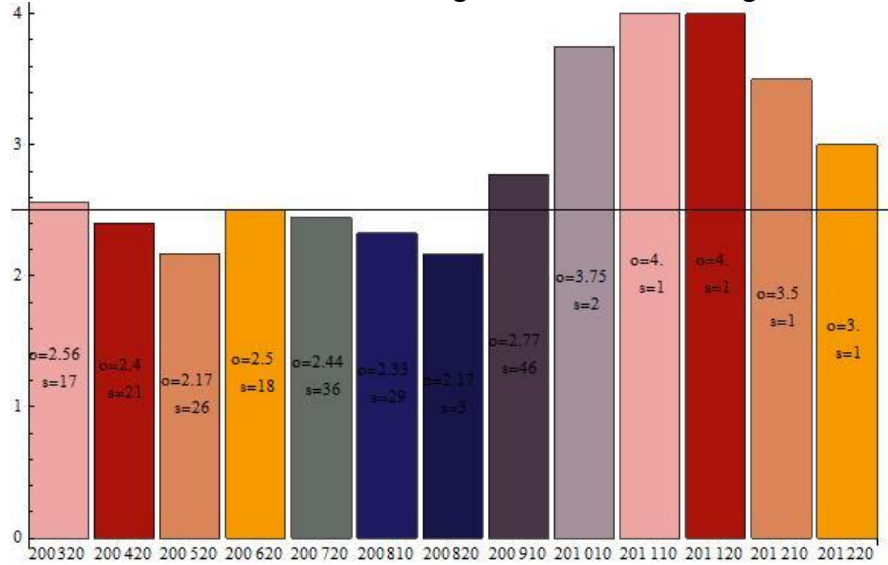
Şekil 4.2: MAT111 dersini alan tüm öğrencilerin dönemlere göre ortalaması



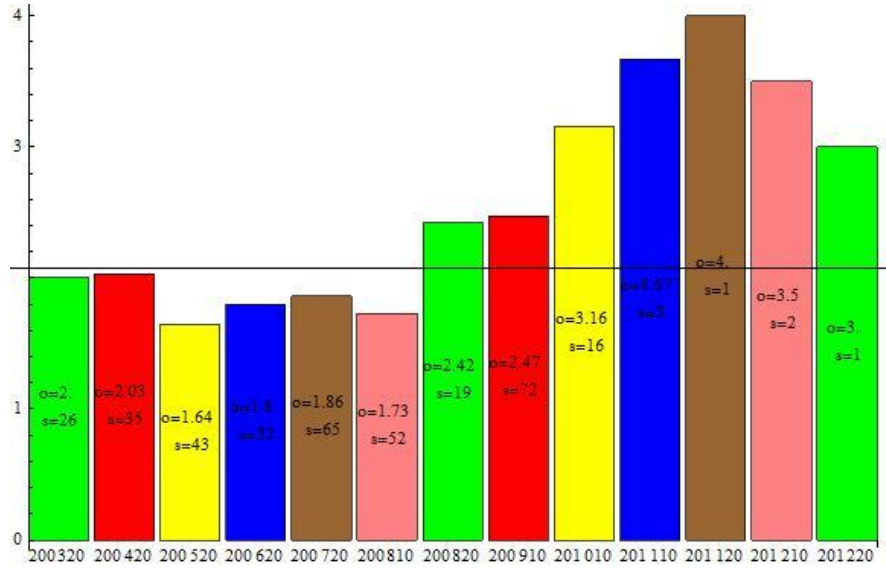
Şekil 4.3: MAT112 dersini ilk kez alanların dönemlere göre ortalaması



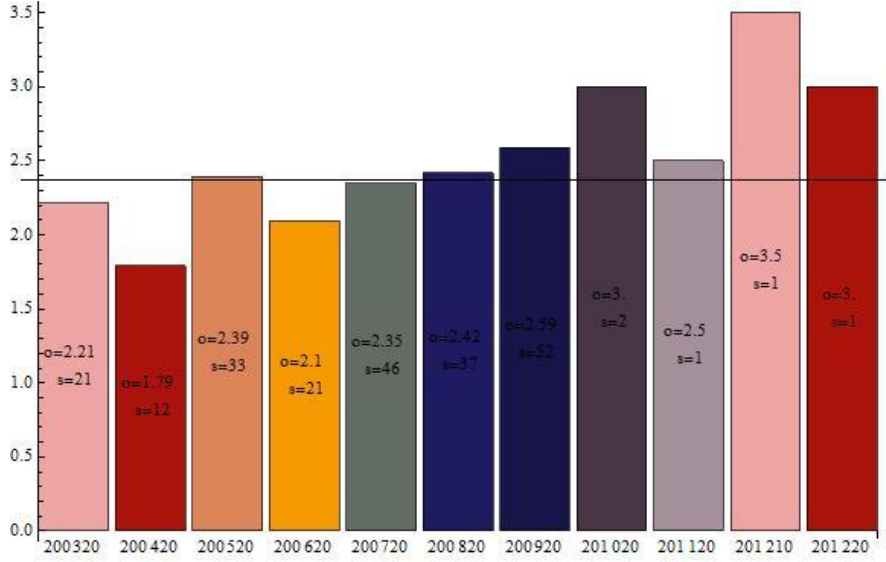
Şekil 4.4: MAT112 dersini alan tüm öğrencilerin dönemlere göre ortalaması



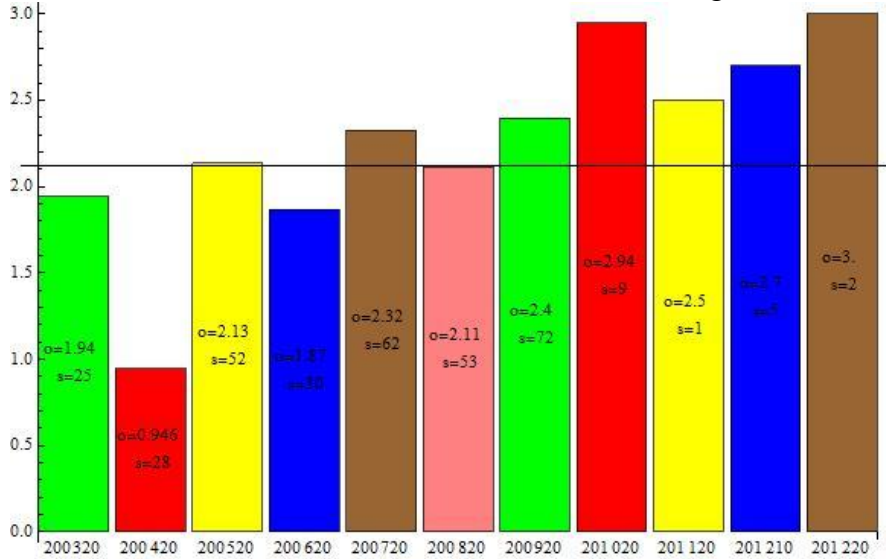
Şekil 4.5: MAT141 dersini ilk kez alanların dönemlere göre ortalaması



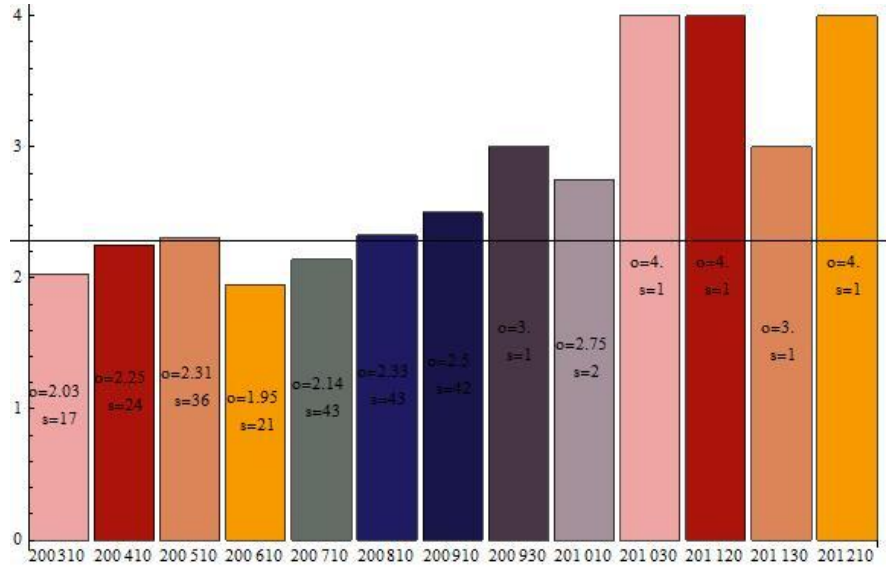
Şekil 4.6: MAT141 dersini alan tüm öğrencilerin dönemlere göre ortalaması



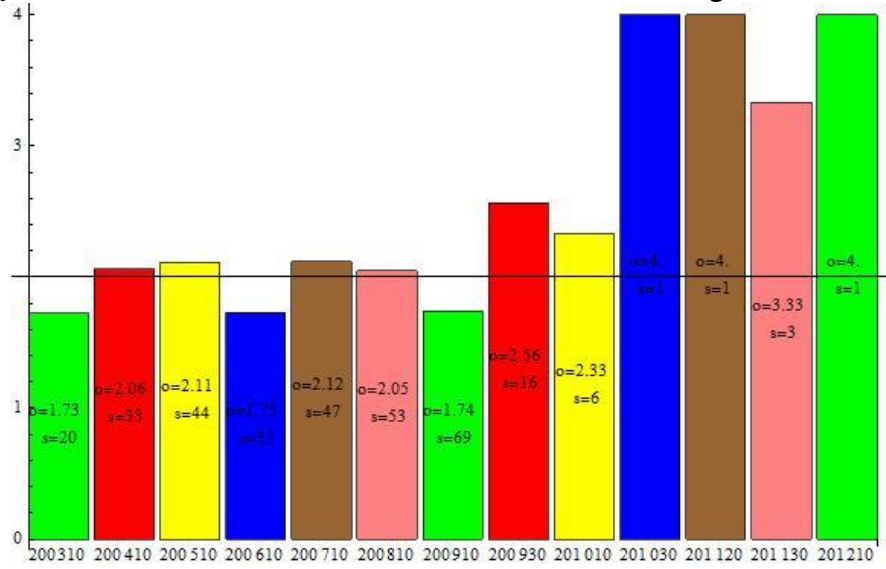
Şekil 4.7: MAT142 dersini ilk kez alanların dönemlere göre ortalaması



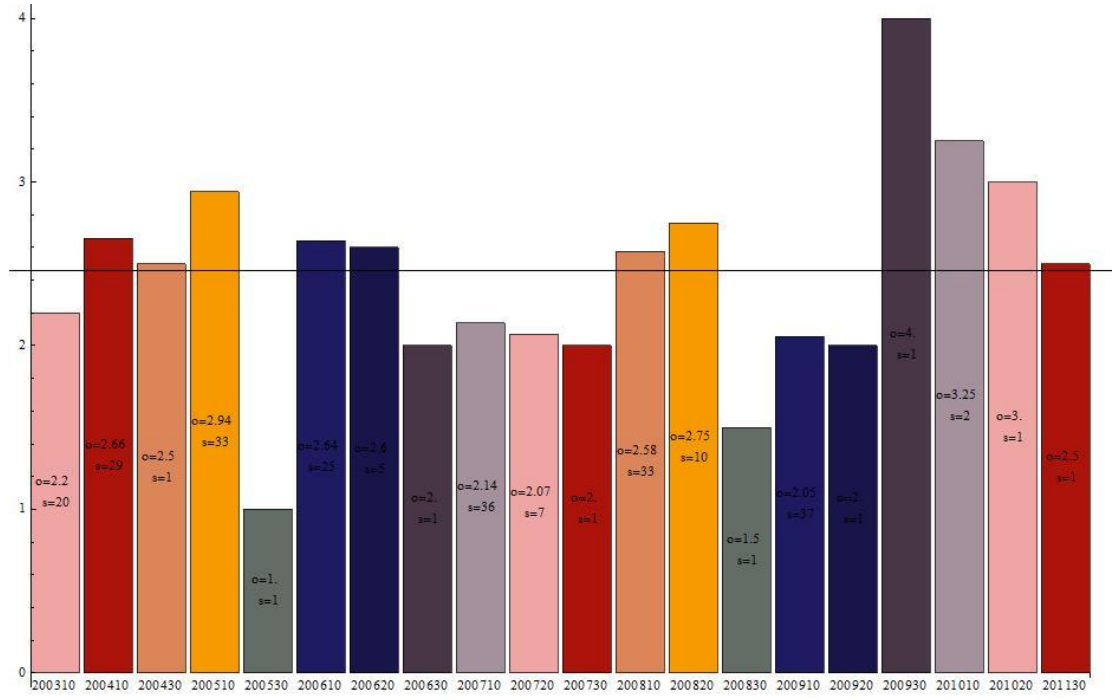
Şekil 4.8: MAT142 dersini alan tüm öğrencilerin dönemlere göre ortalaması



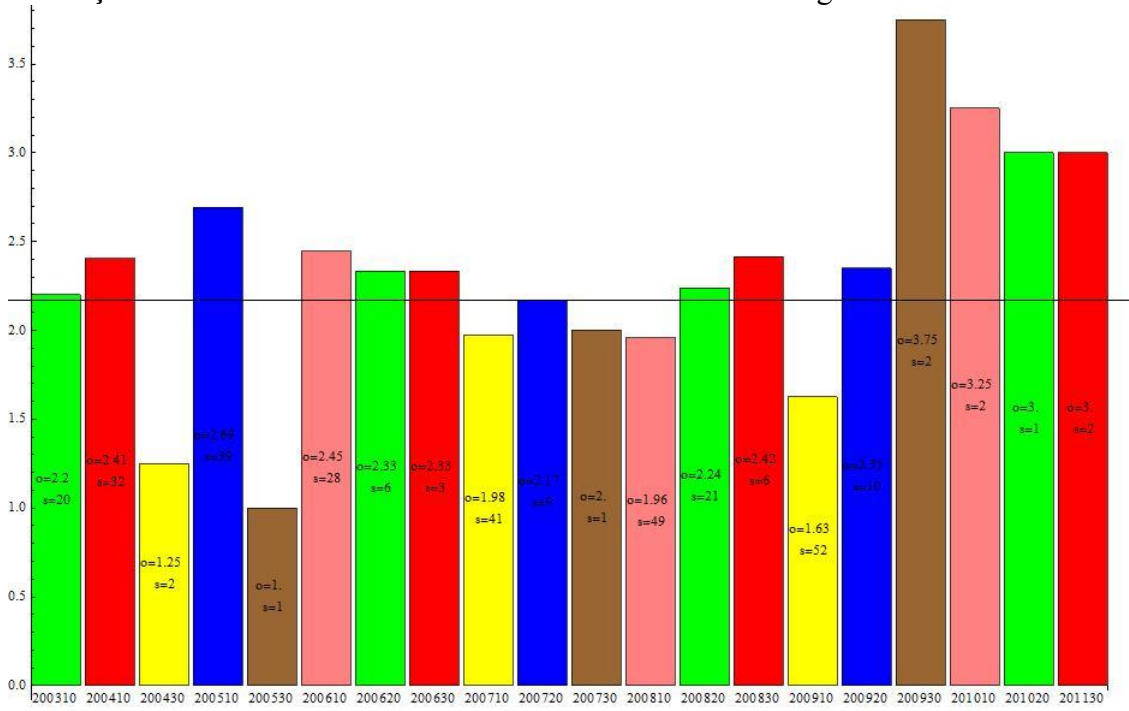
Şekil 4.9: MAT175 dersini ilk kez alanların dönemlere göre ortalaması



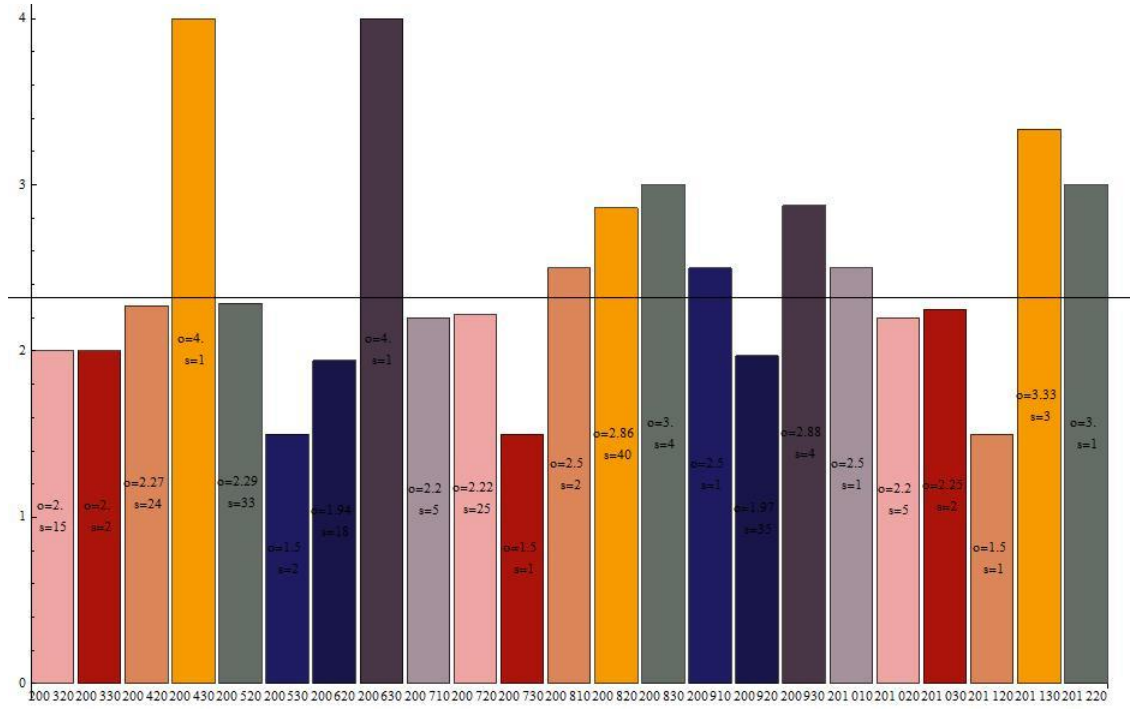
Şekil 4.10: MAT175 dersini alan tüm öğrencilerin dönemlere göre ortalaması



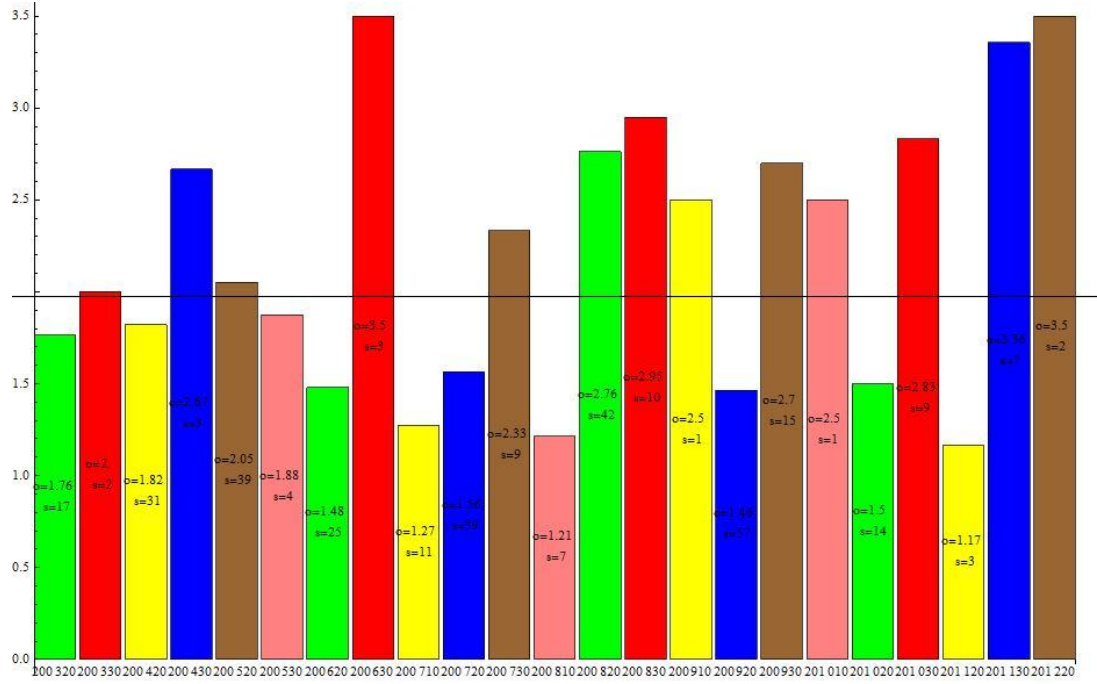
Şekil 4.11: FIZ101E dersini ilk kez alanların dönemlere göre ortalaması



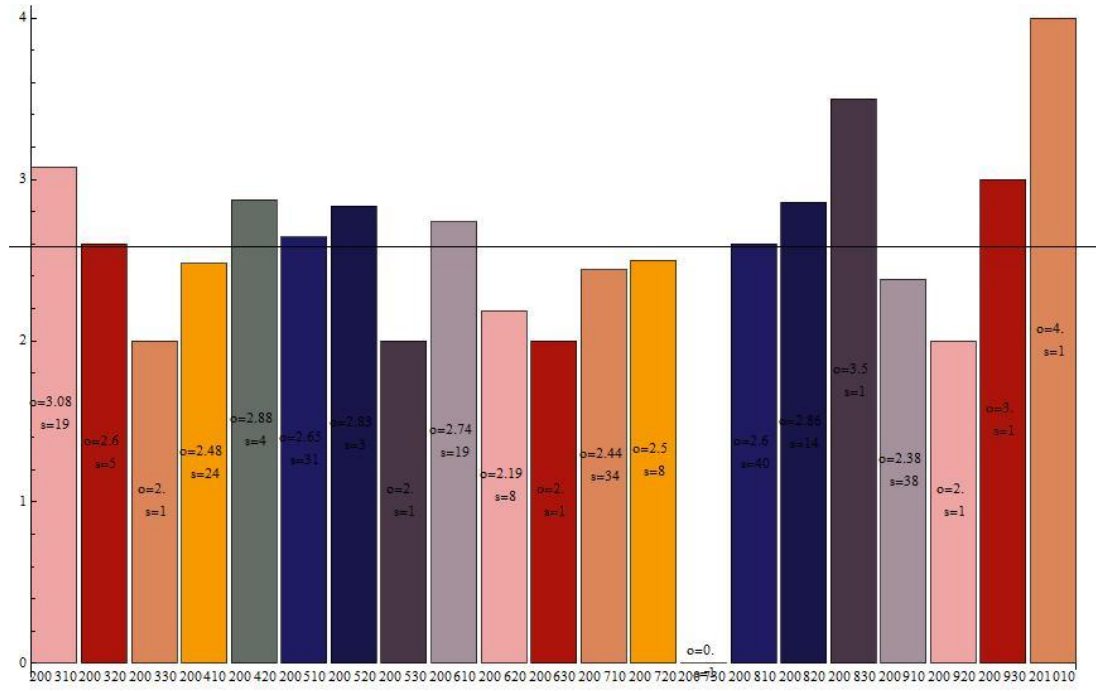
Şekil 4.12: FIZ101E dersini alan tüm öğrencilerin dönemlere göre ortalaması



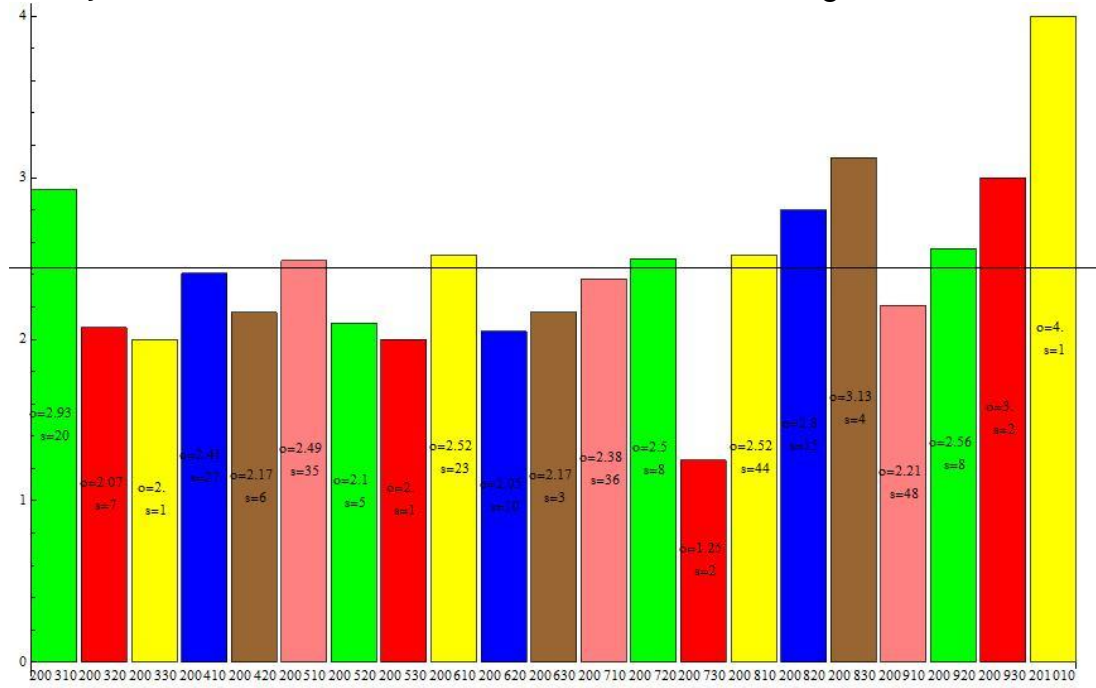
Şekil 4.13: FIZ106 dersini ilk kez alanların dönemlere göre ortalaması



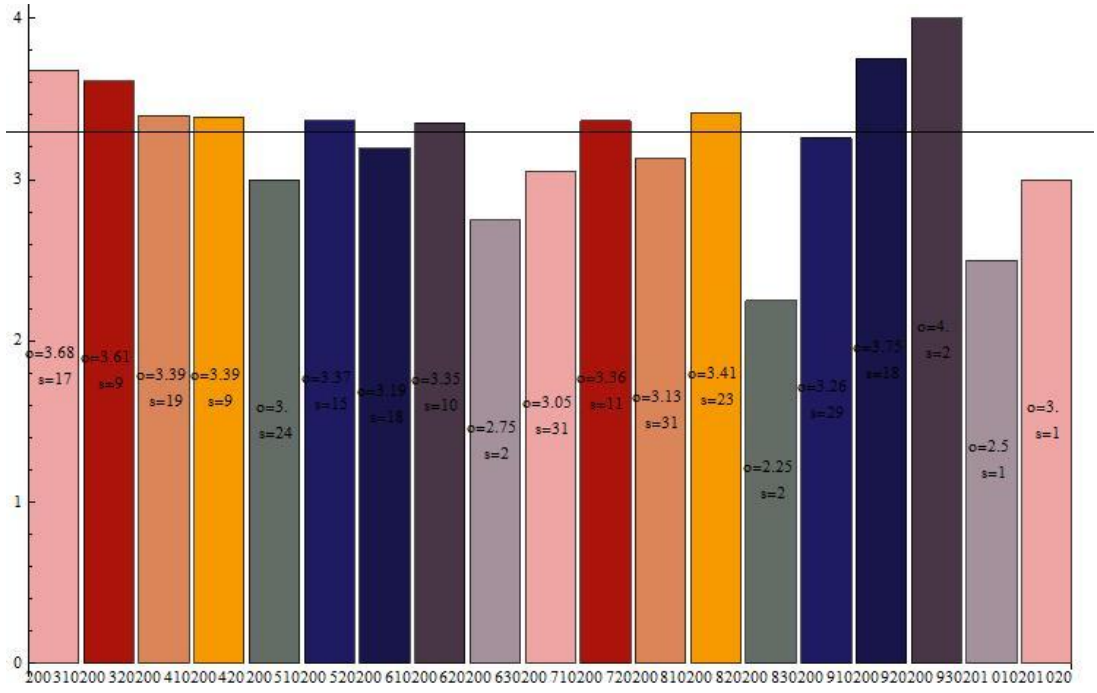
Şekil 4.14: FIZ106 dersini alan tüm öğrencilerin dönemlere göre ortalaması



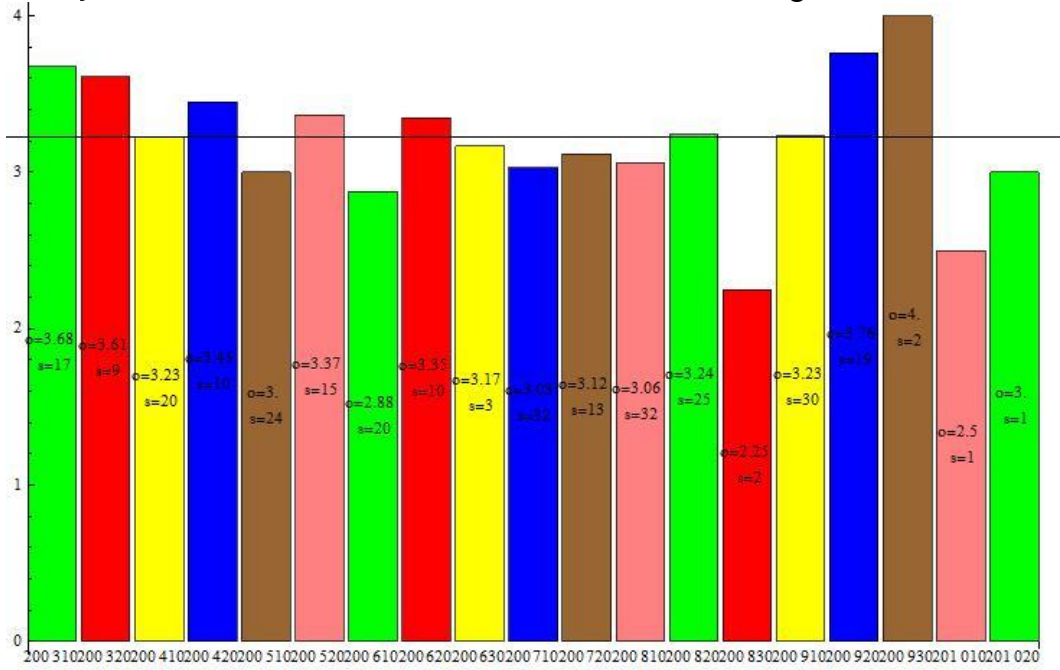
Şekil 4.15: KIM101E dersini ilk kez alanların dönemlere göre ortalaması



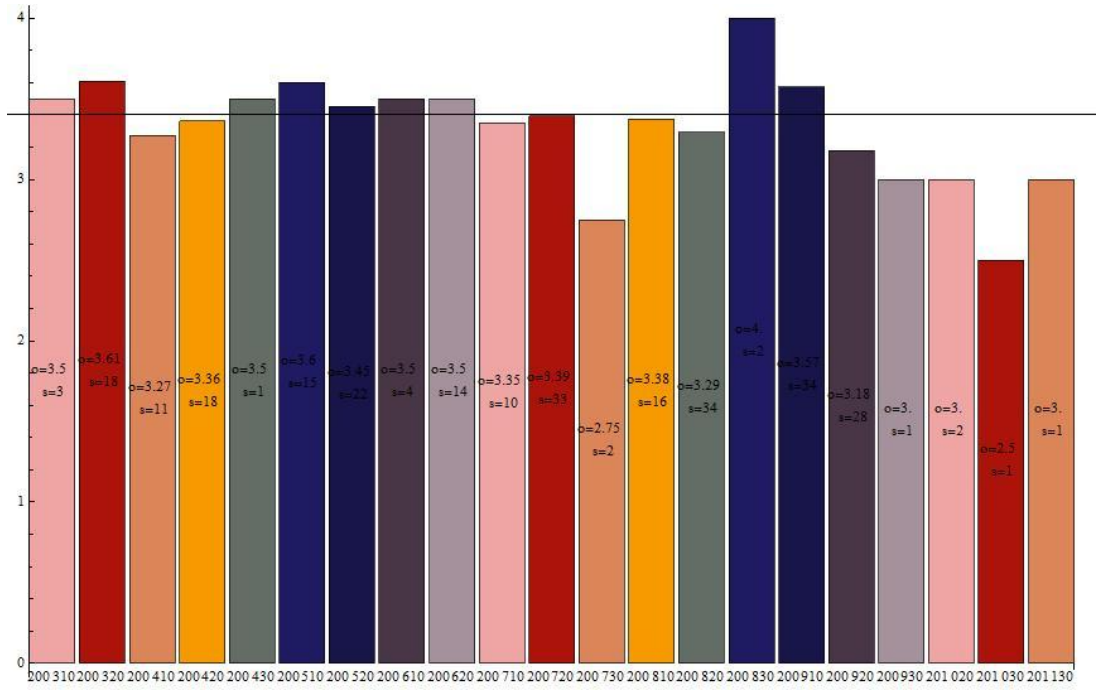
Şekil 4.16: KIM101E dersini alan tüm öğrencilerin dönemlere göre ortalaması



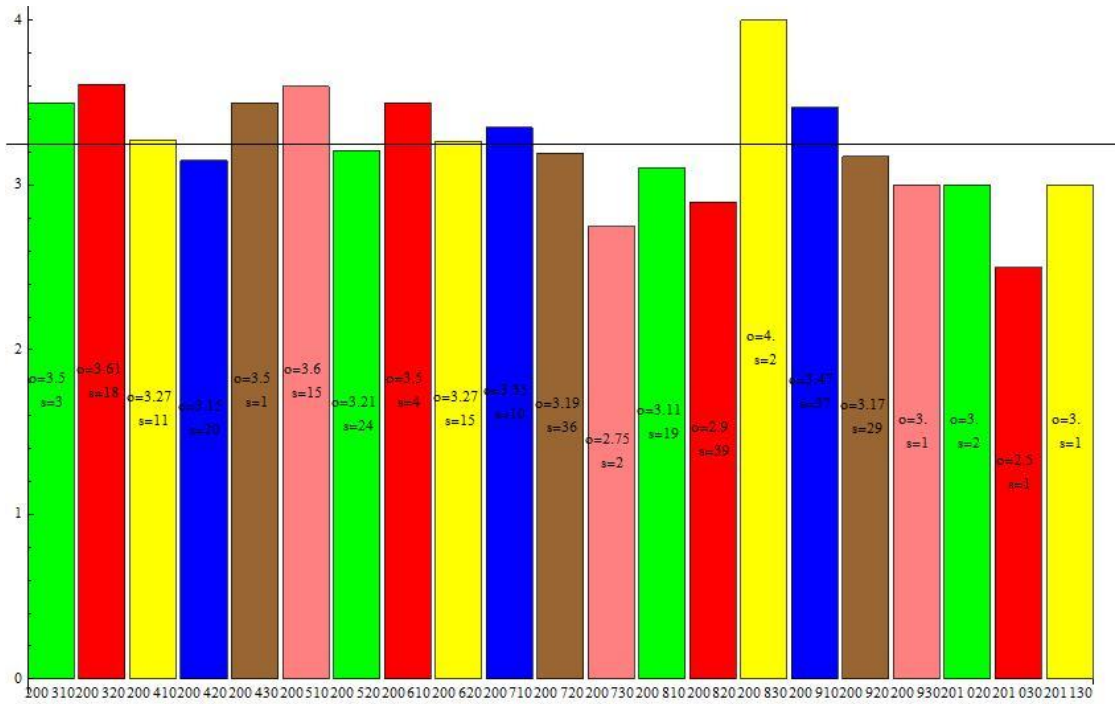
Şekil 4.17: ING101 dersini ilk kez alanların dönemlere göre ortalaması



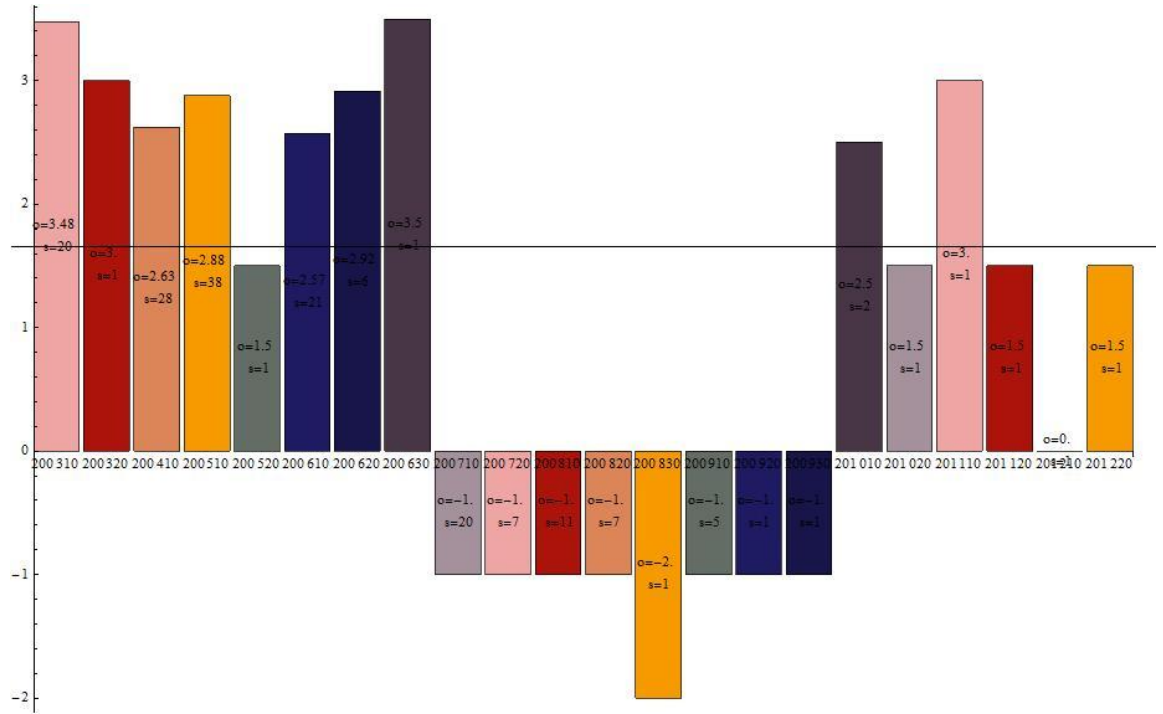
Şekil 4.18: ING101 dersini alan tüm öğrencilerin dönemlere göre ortalaması



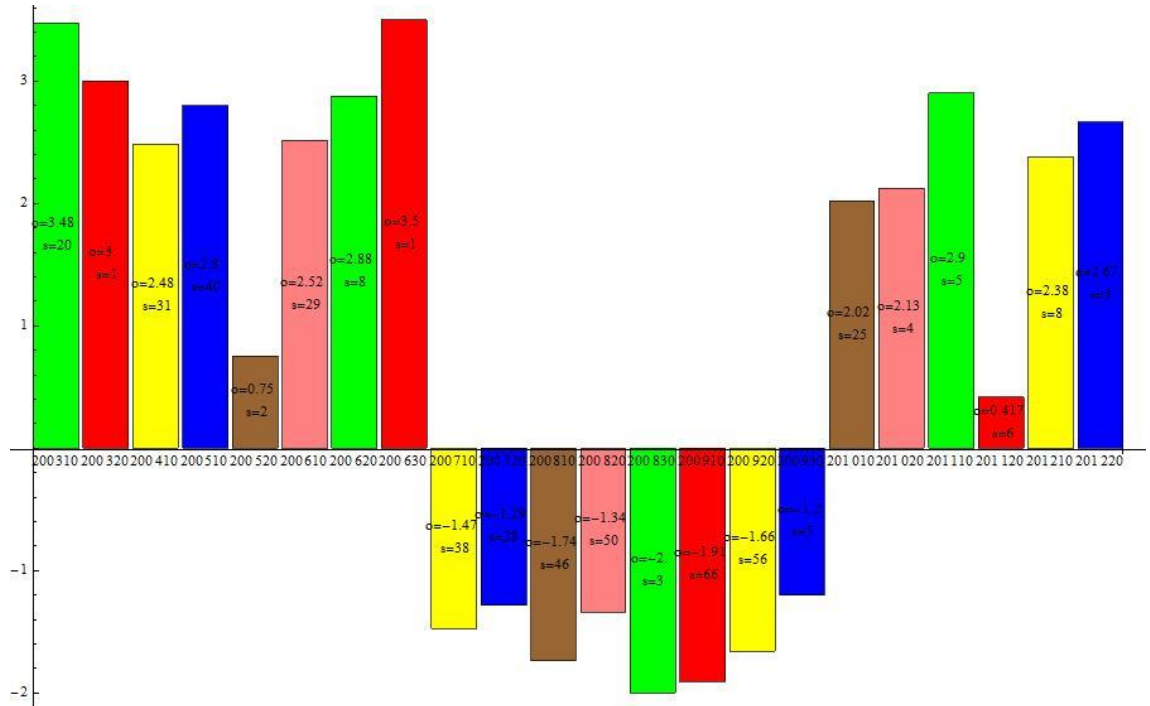
Şekil 4.19: ING102 dersini ilk kez alanların dönemlere göre ortalaması



Şekil 4.20: ING102 dersini alan tüm öğrencilerin dönemlere göre ortalaması

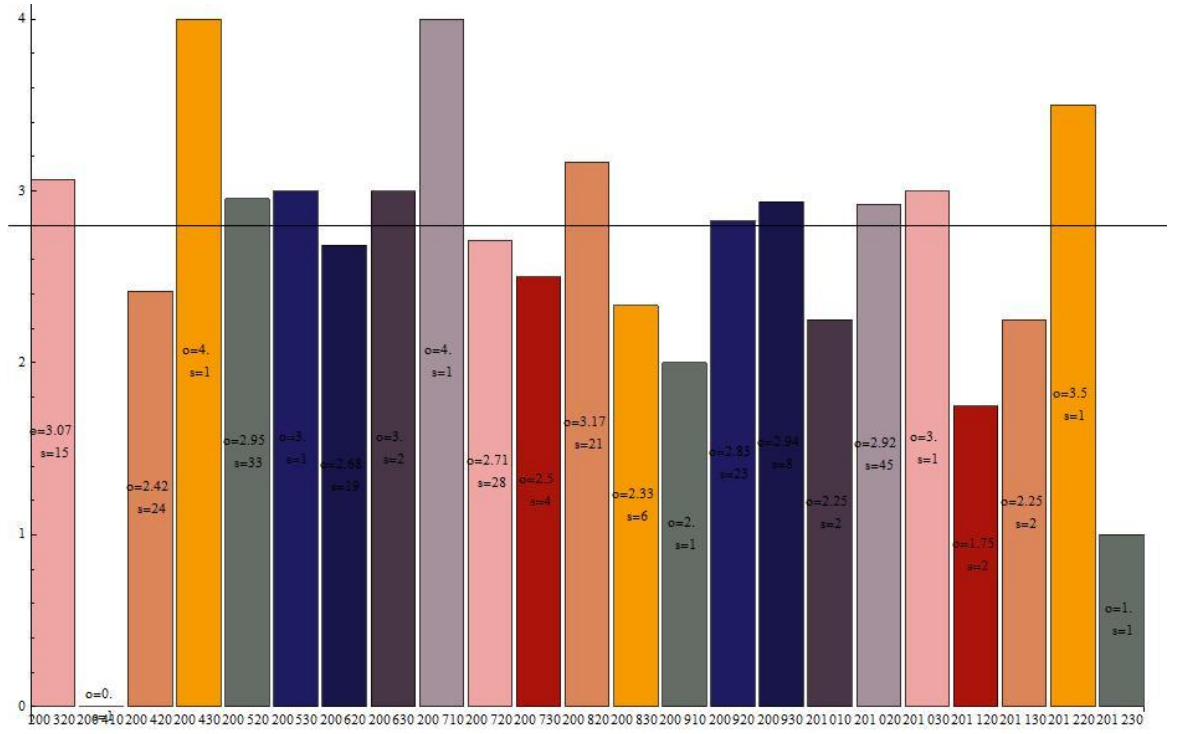


Şekil 4.21: BIL101E dersini ilk kez alanların dönemlere göre ortalaması*

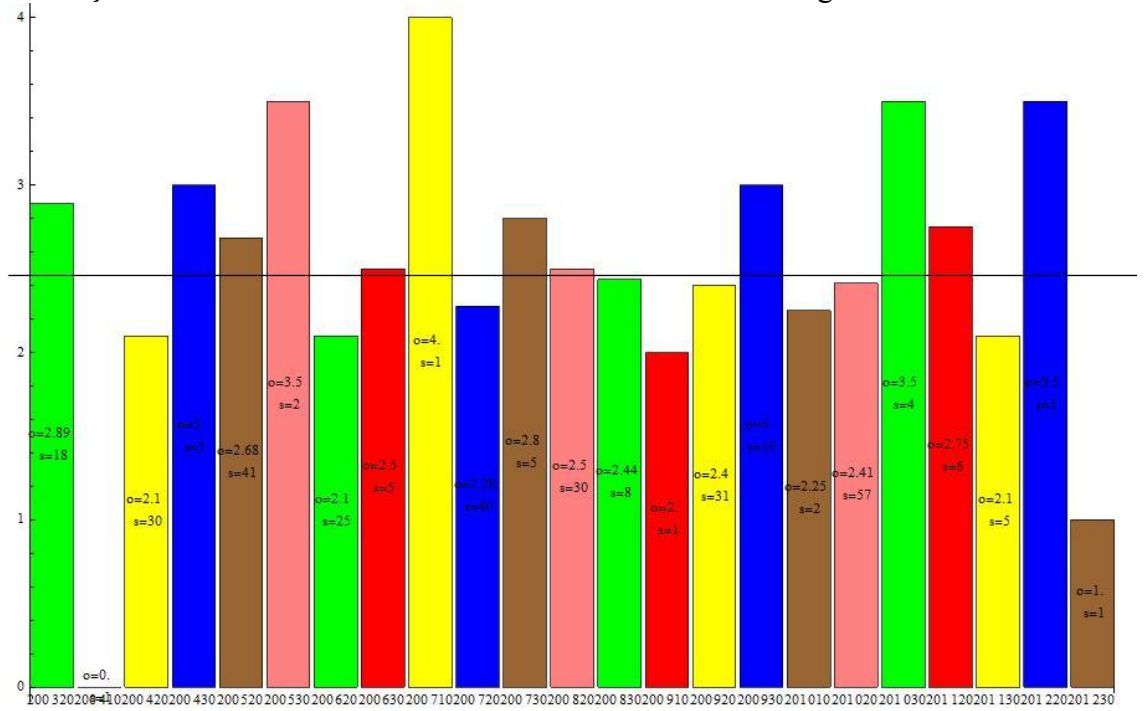


Şekil 4.22: BIL101E dersini alan tüm öğrencilerin dönemlere göre ortalaması*

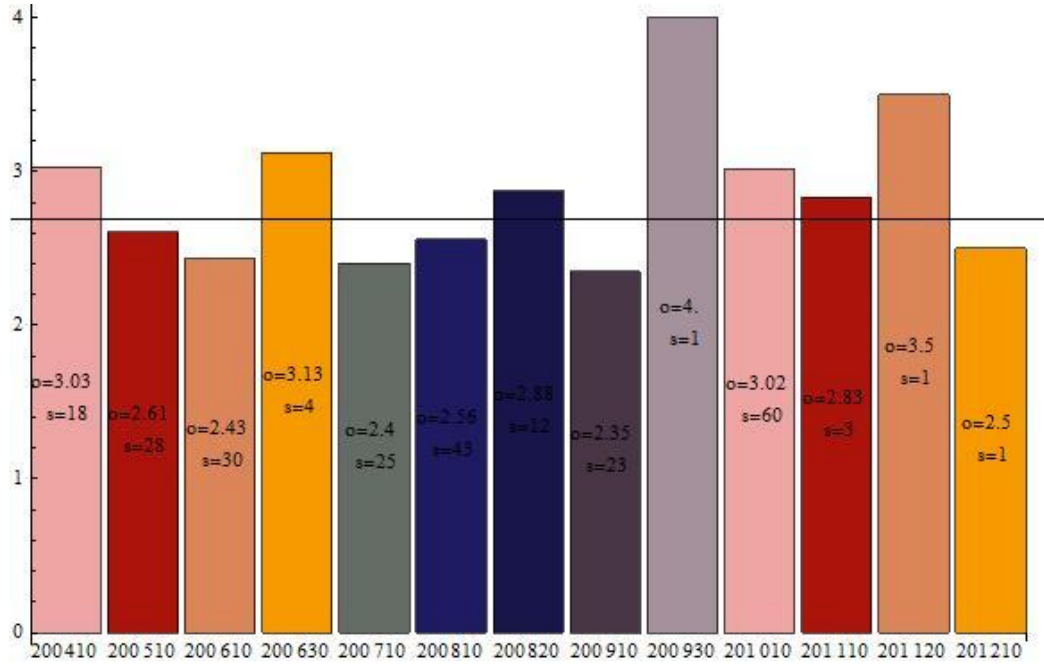
*Ortalamaların negatif değer aldığı dönemlerde BIL101E dersinin not karşılığı olmayan “BL” “BZ” harfleri ile değerlendirildiği görülmektedir. Bu dönemler için de ortalama bilgisi elde etmek için “BL” için “-1”, “BZ” için “-2” değeri verilerek hesap yapılmıştır.



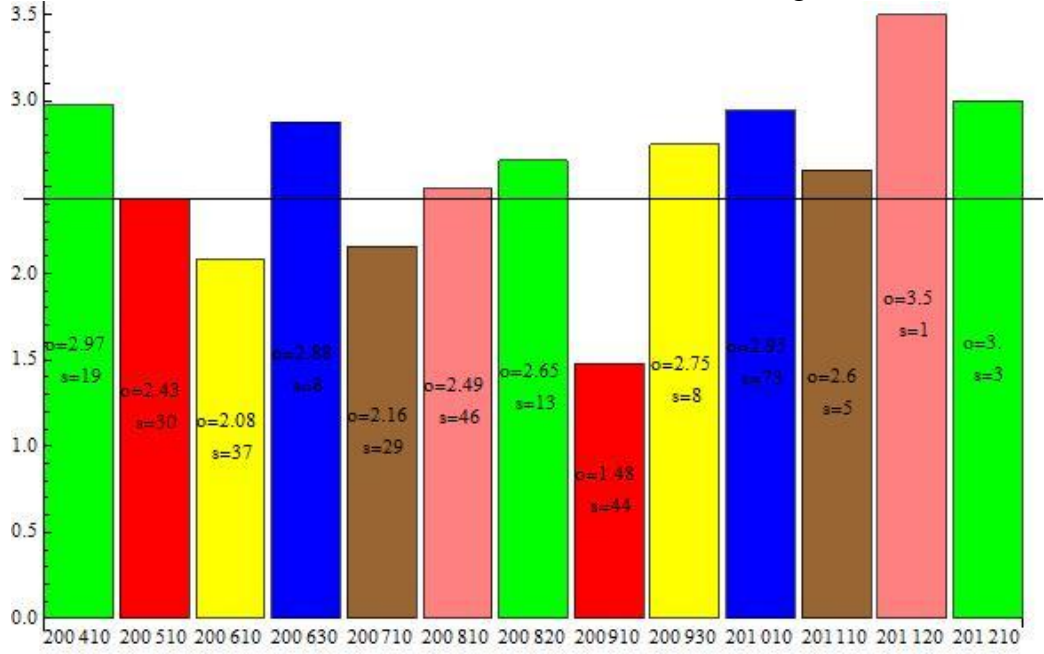
Şekil 4.23: BIL104E dersini ilk kez alanların dönemlere göre ortalaması



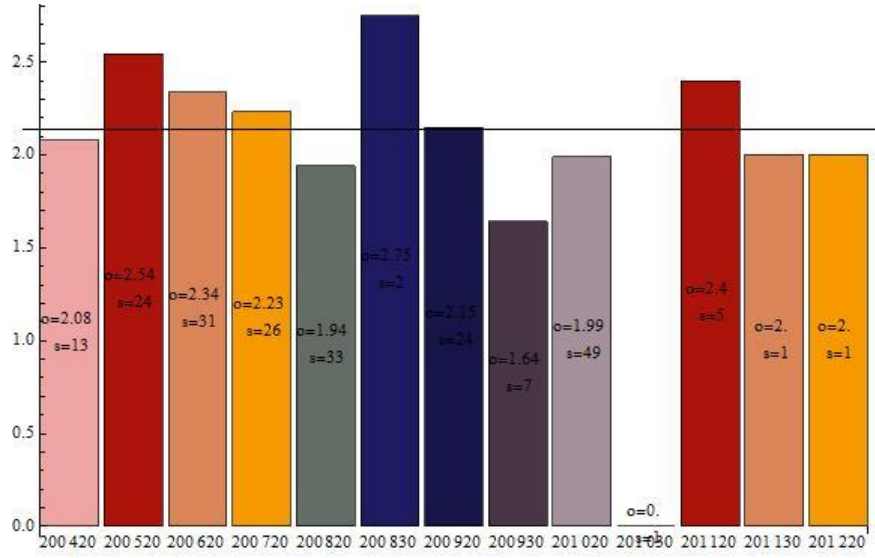
Şekil 4.24: BIL104E dersini alan tüm öğrencilerin dönemlere göre ortalaması



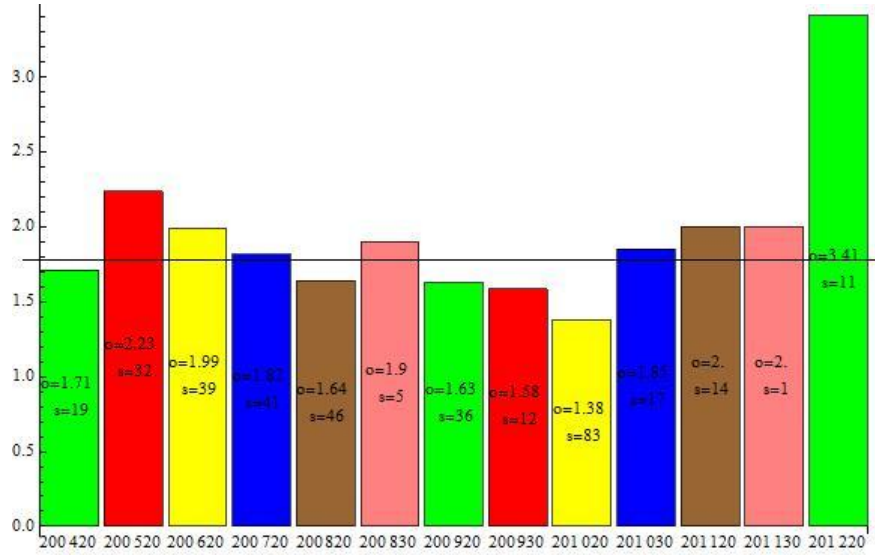
Şekil 4.25: MAT213 dersini ilk kez alanların dönemlere göre ortalaması



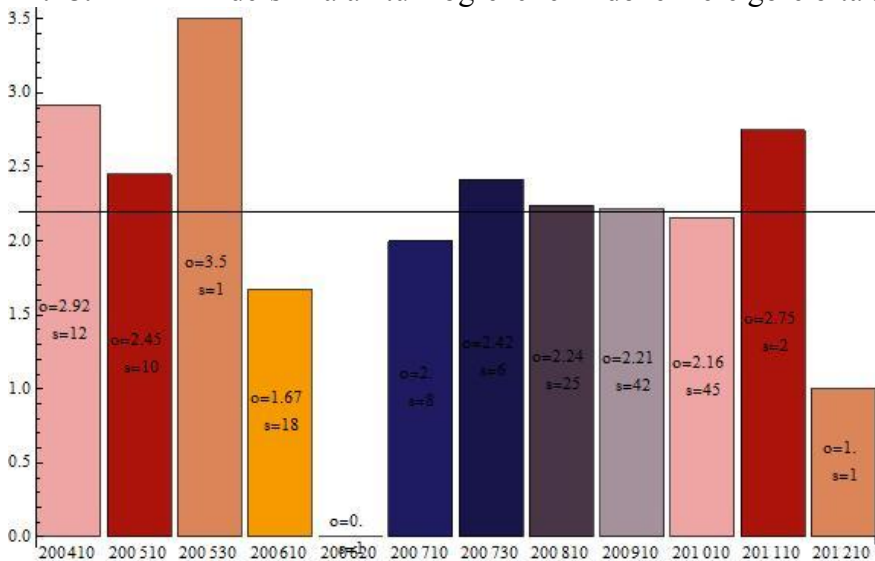
Şekil 4.26: MAT213 dersini alan tüm öğrencilerin dönemlere göre ortalaması



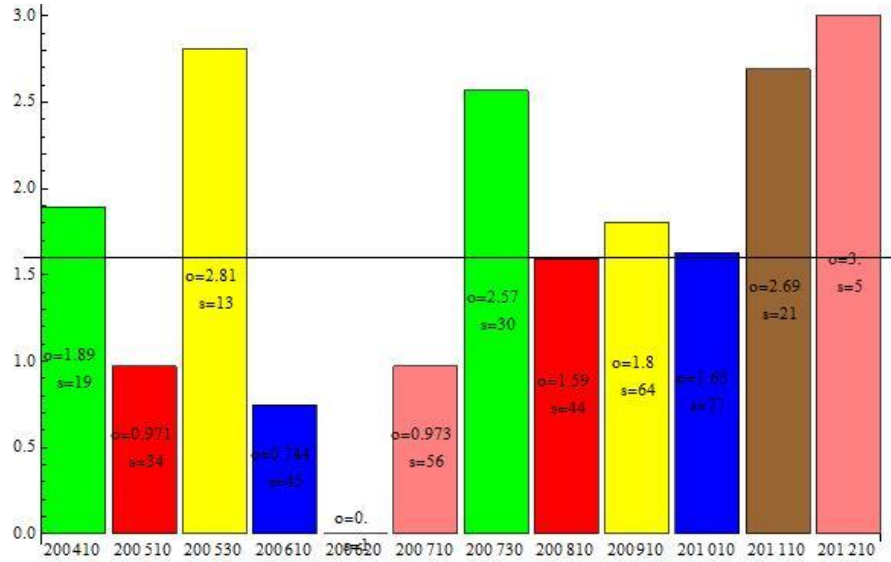
Şekil 4.27: MAT214 dersini ilk kez alanların dönemlere göre ortalaması



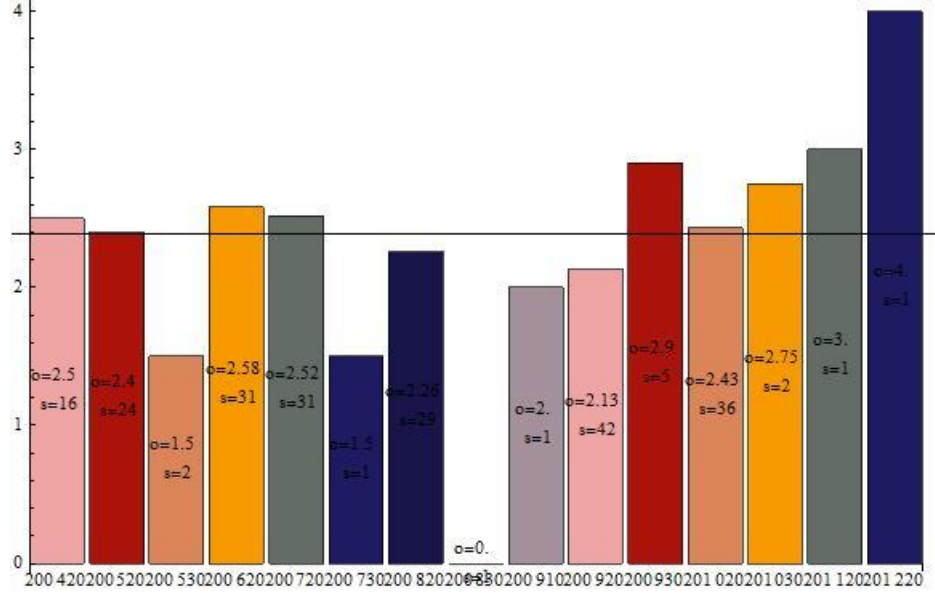
Şekil 4.28: MAT214 dersini alan tüm öğrencilerin dönemlere göre ortalaması



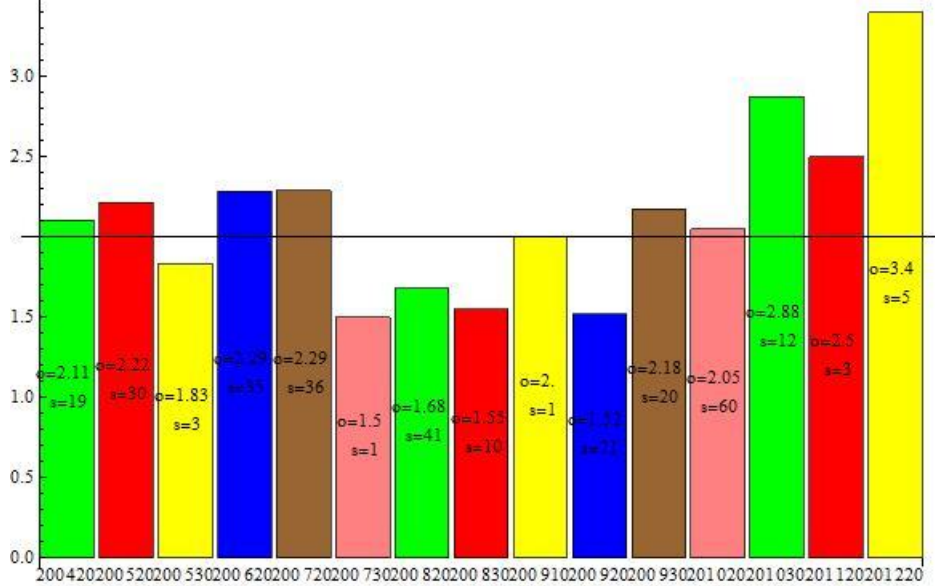
Şekil 4.29: MAT221 dersini ilk kez alanların dönemlere göre ortalaması



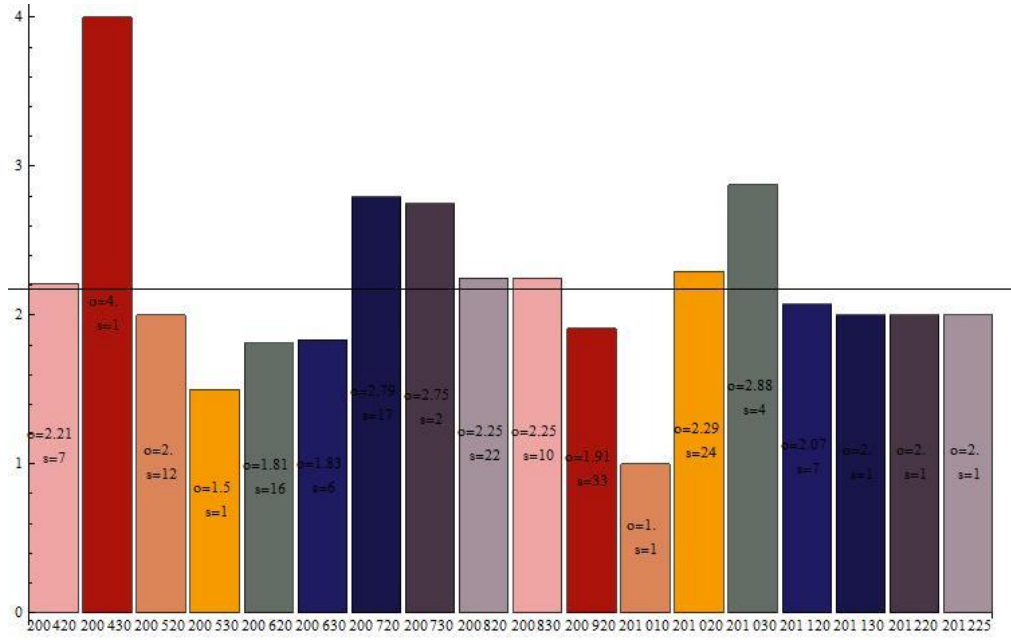
Şekil 4.30: MAT221 dersini alan tüm öğrencilerin dönemlere göre ortalaması



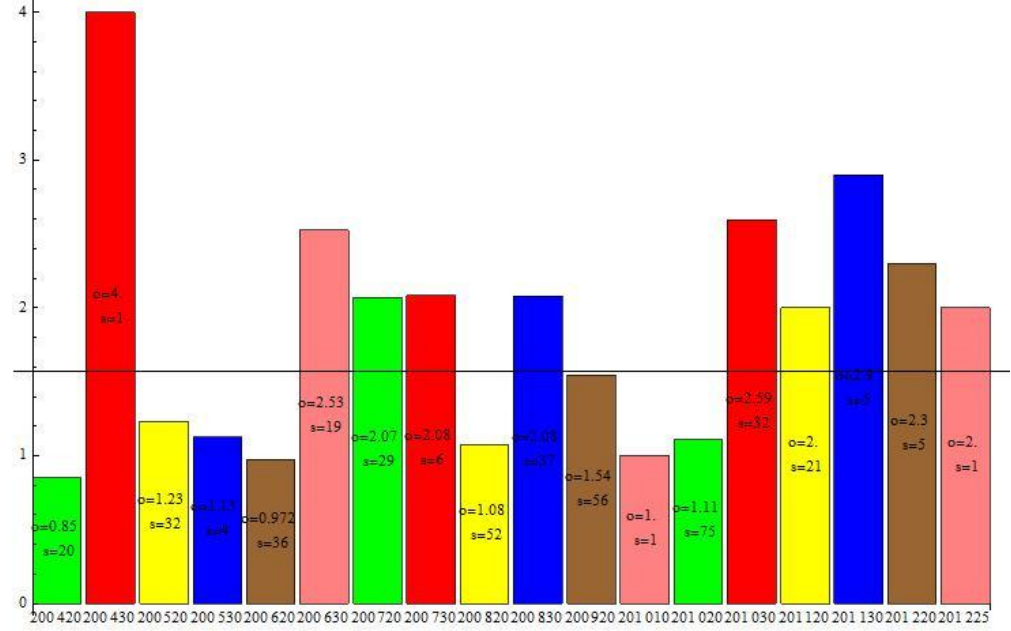
Şekil 4.31: MAT232 dersini ilk kez alanların dönemlere göre ortalaması



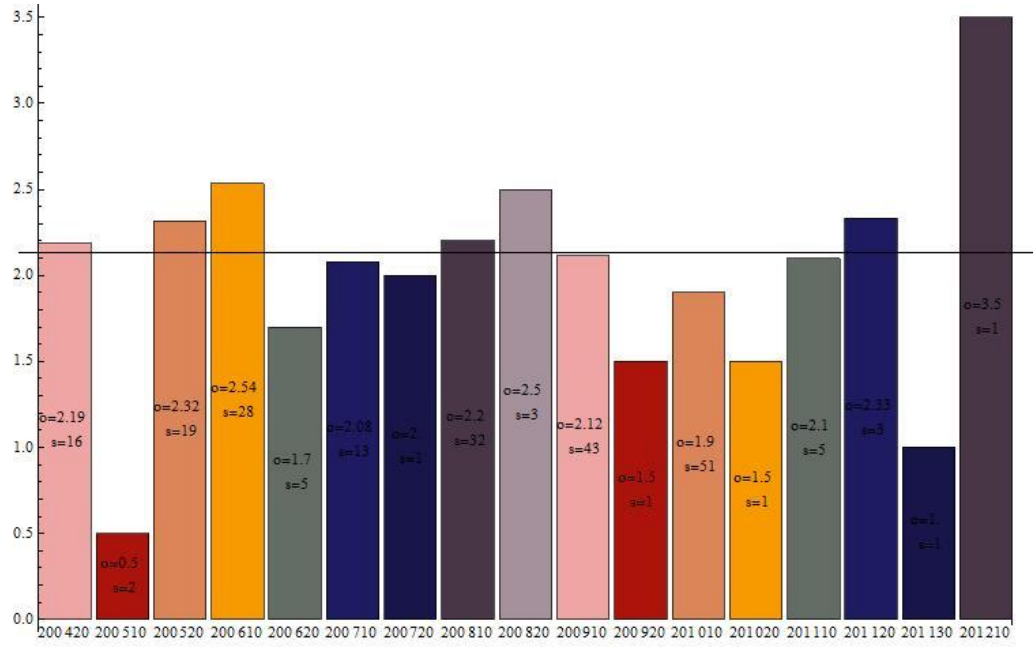
Şekil 4.32: MAT232 dersini alan tüm öğrencilerin dönemlere göre ortalaması



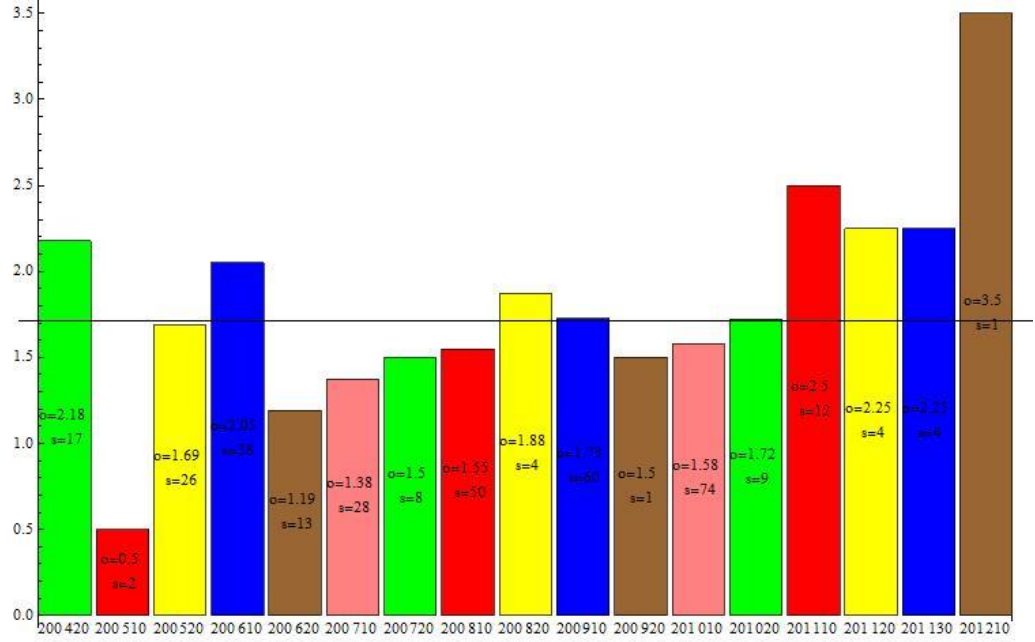
Şekil 4.33: MAT242 dersini ilk kez alanların dönemlere göre ortalaması



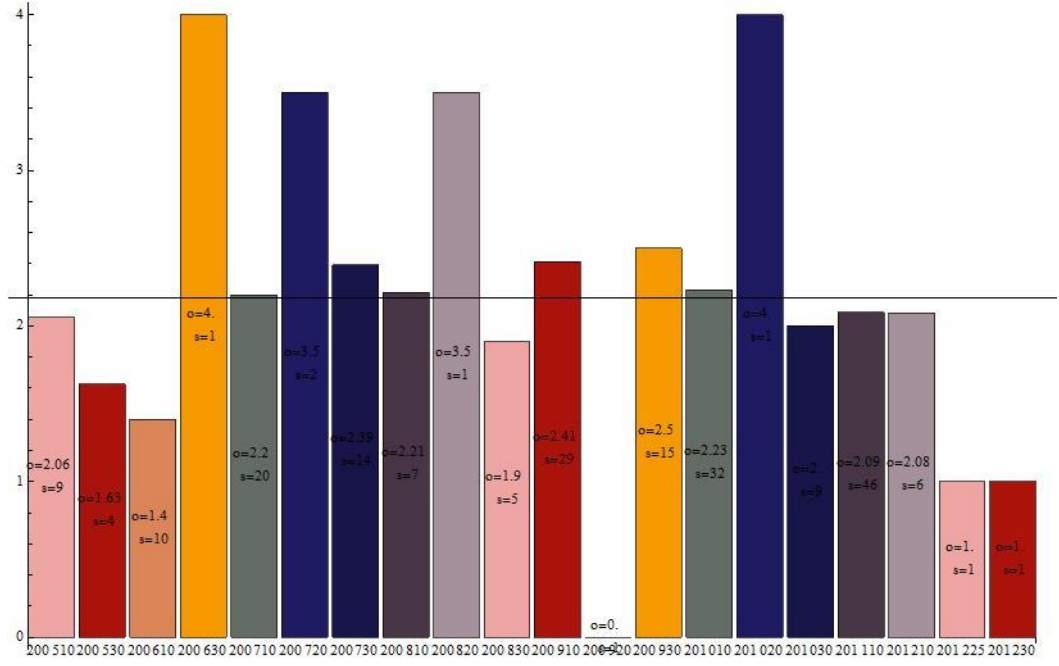
Şekil 4.34: MAT242 dersini alan tüm öğrencilerin dönemlere göre ortalaması



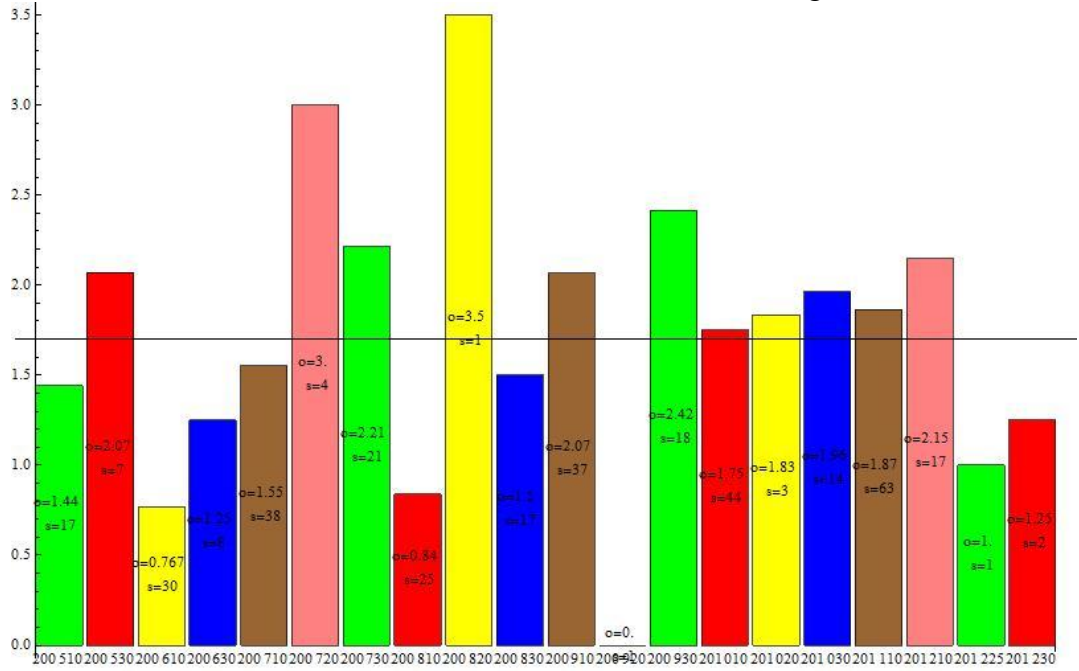
Şekil 4.35: MAT322E dersini ilk kez alanların dönemlere göre ortalaması



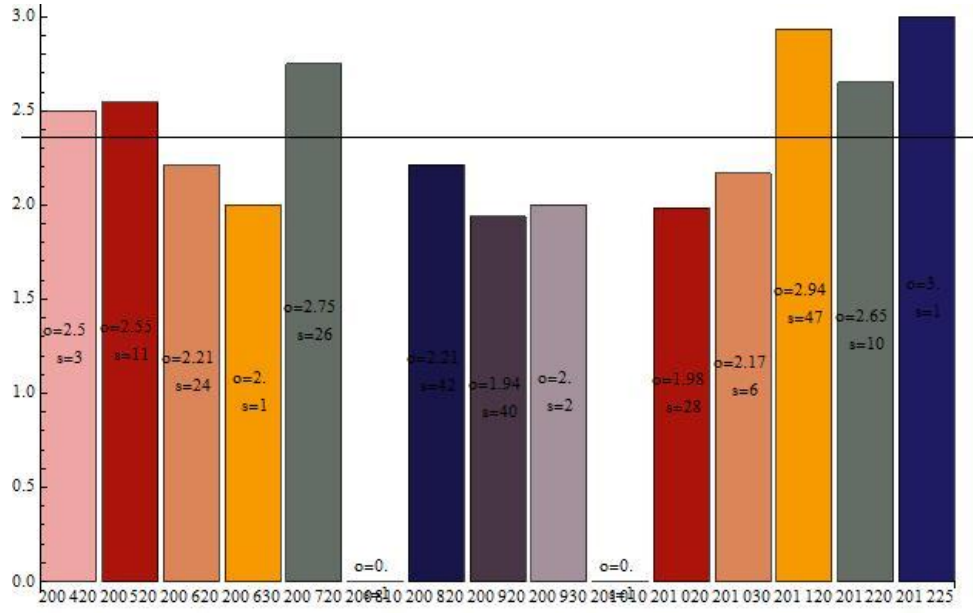
Şekil 4.36: MAT322E dersini alan tüm öğrencilerin dönemlere göre ortalaması



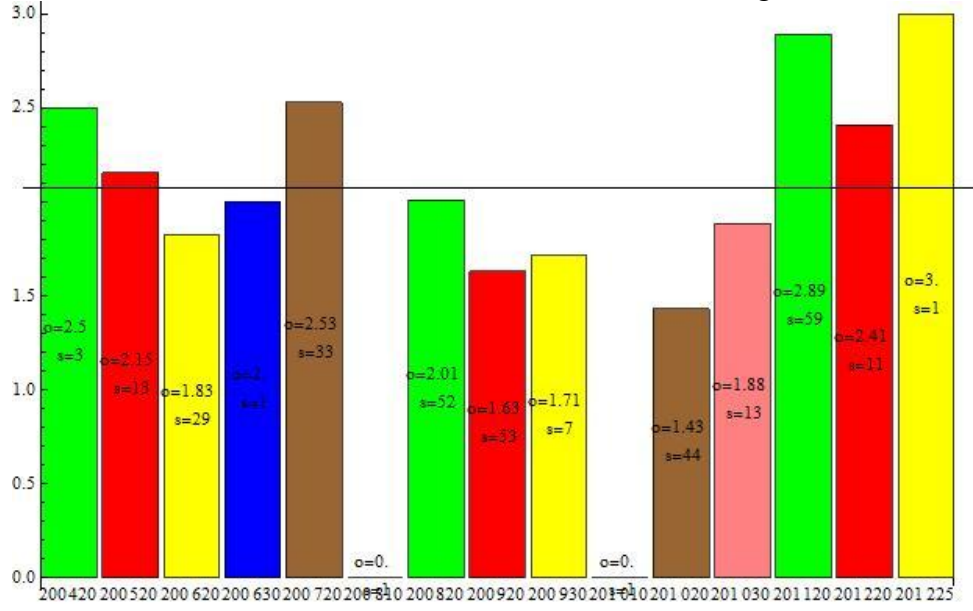
Şekil 4.37: MAT331 dersini ilk kez alanların dönemlere göre ortalaması



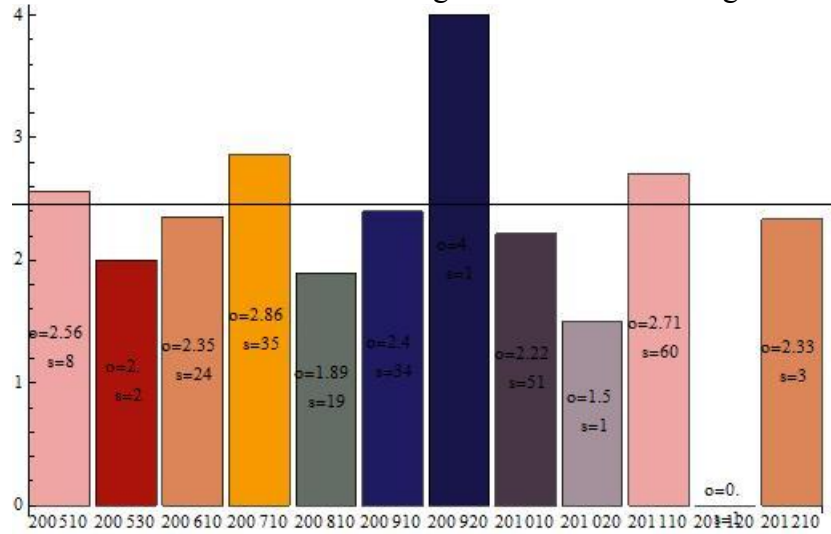
Şekil 4.38: MAT331 dersini alan tüm öğrencilerin dönemlere göre ortalaması



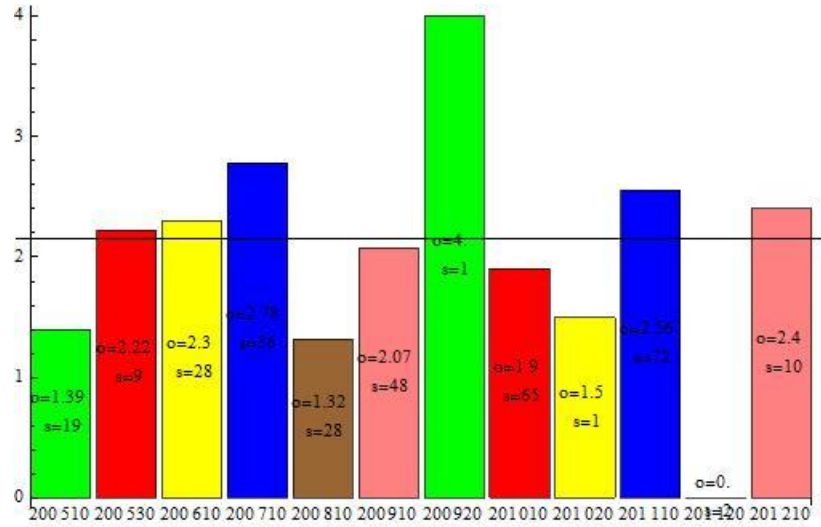
Şekil 4.39: MAT332 dersini ilk kez alanların dönemlere göre ortalaması



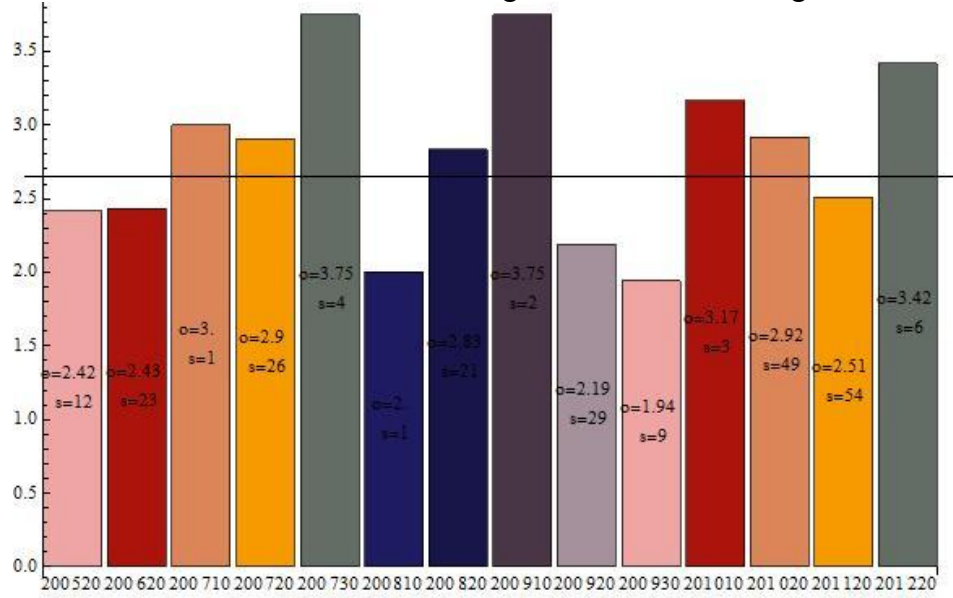
Şekil 4.40: MAT332 dersini alan tüm öğrencilerin dönemlere göre ortalaması



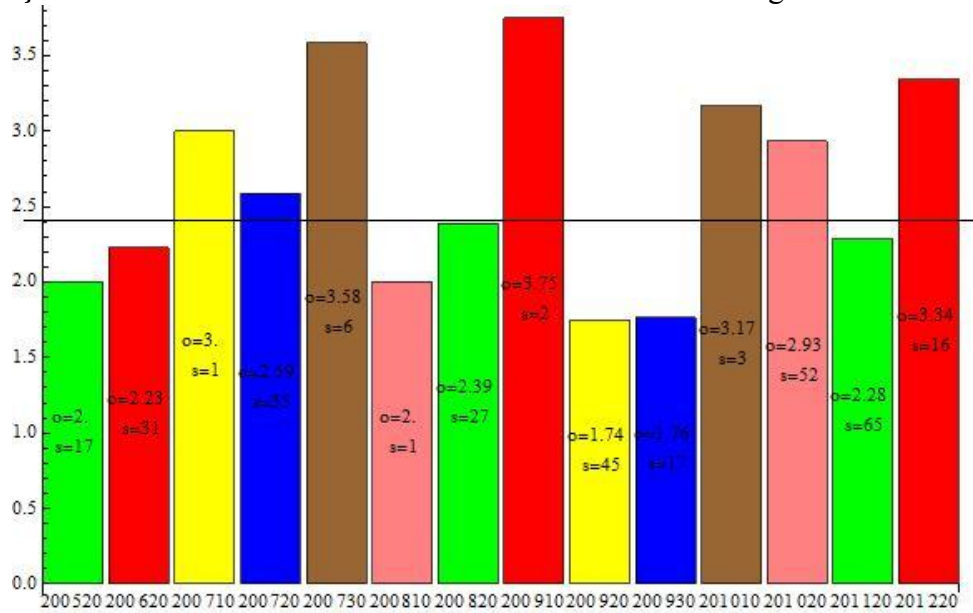
Şekil 4.41: MAT341E dersini ilk kez alanların dönemlere göre ortalaması



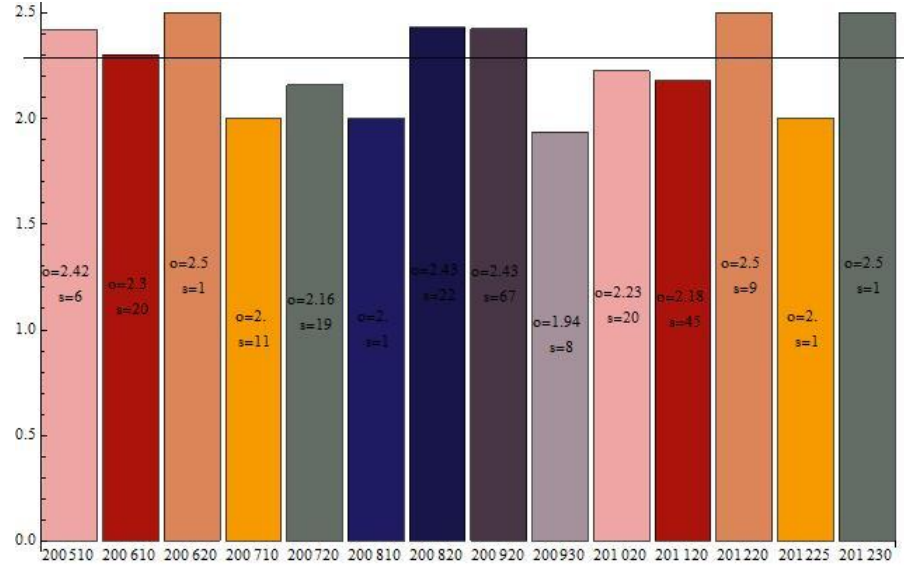
Şekil 4.42: MAT341E dersini alan tüm öğrencilerin dönemlere göre ortalaması



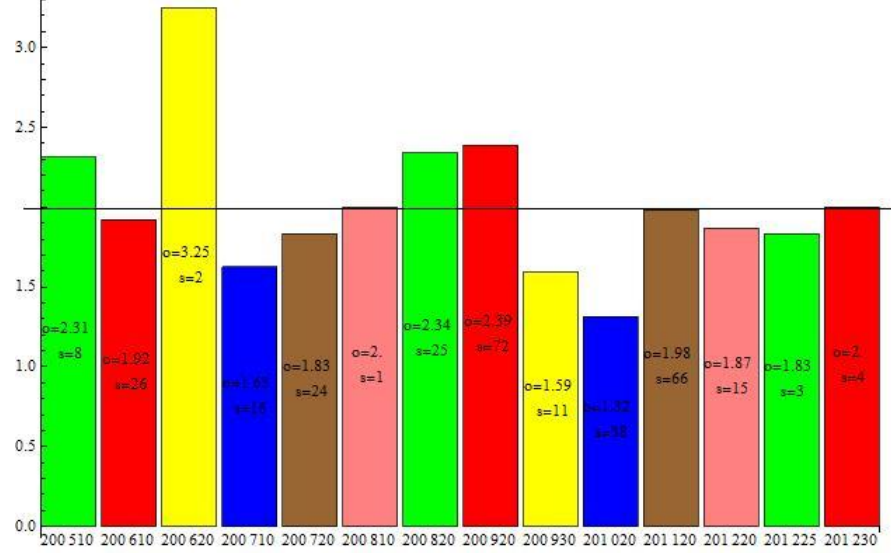
Şekil 4.43: MAT342E dersini ilk kez alanların dönemlere göre ortalaması



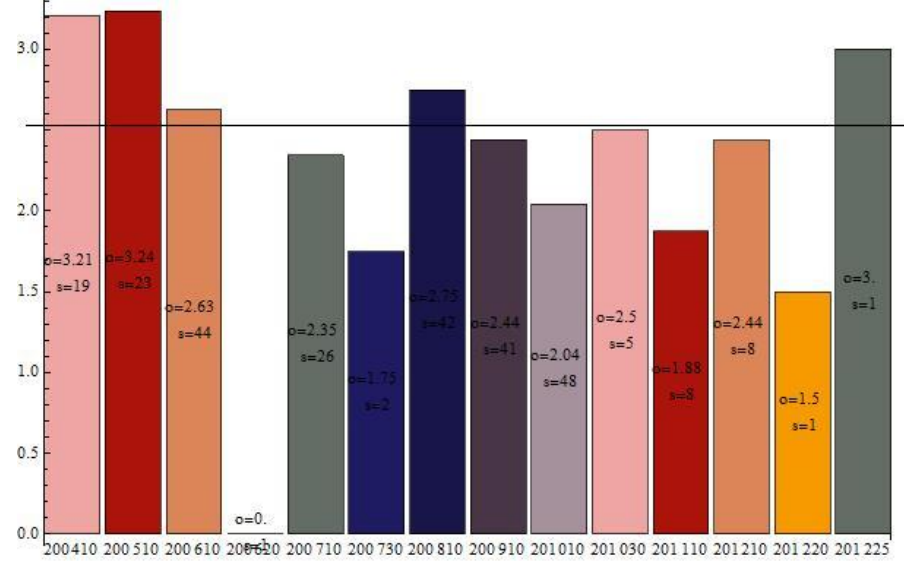
Şekil 4.44: MAT342E dersini alan tüm öğrencilerin dönemlere göre ortalaması



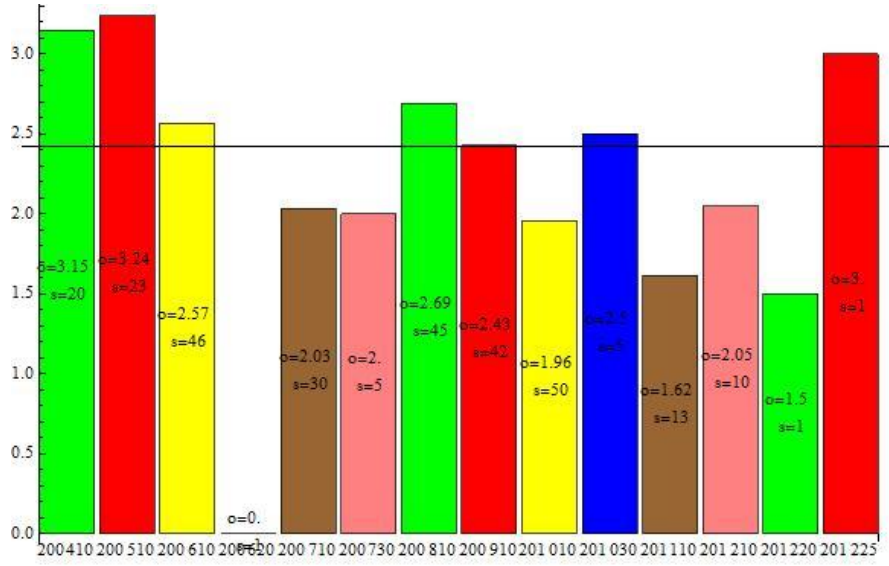
Şekil 4.45: MAT355 dersini ilk kez alanların dönemlere göre ortalaması



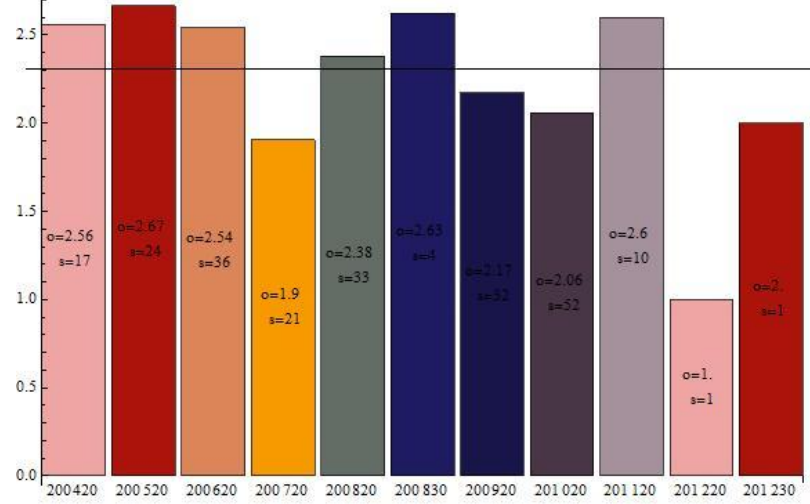
Şekil 4.46: MAT355 dersini alan tüm öğrencilerin dönemlere göre ortalaması



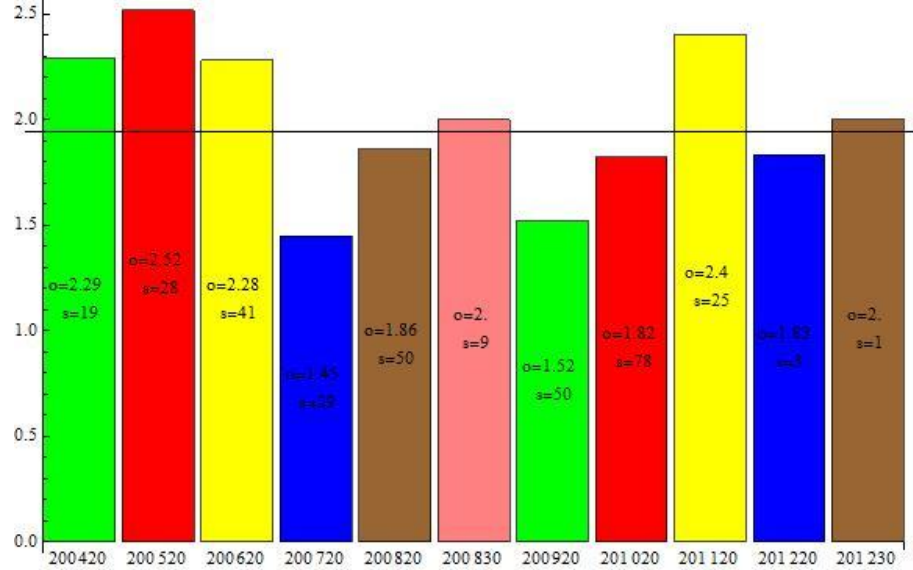
Şekil 4.47: MUH211 dersini ilk kez alanların dönemlere göre ortalaması



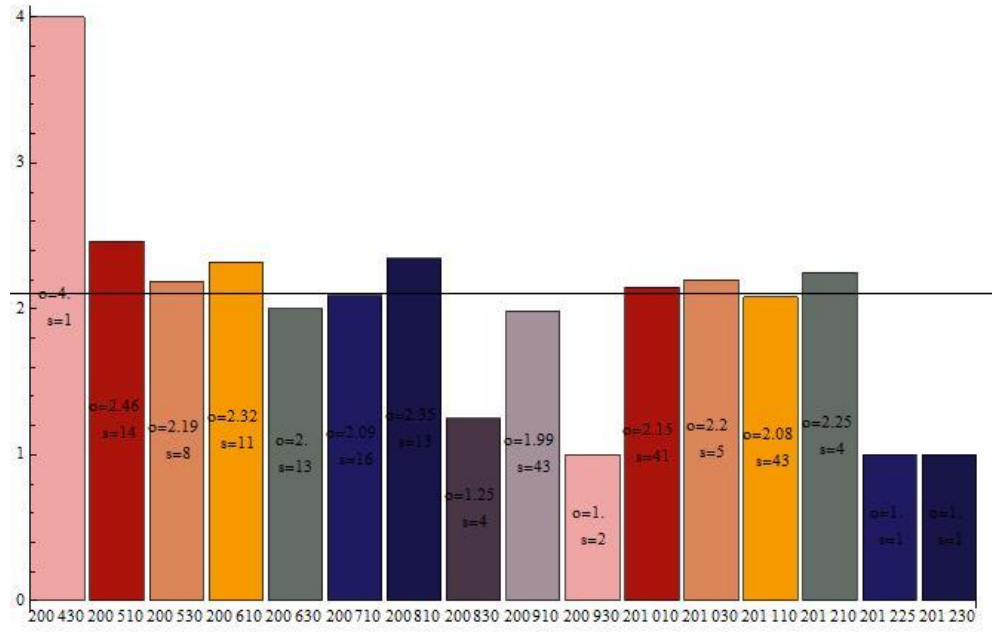
Şekil 4.48: MUH211 dersini alan tüm öğrencilerin dönemlere göre ortalaması



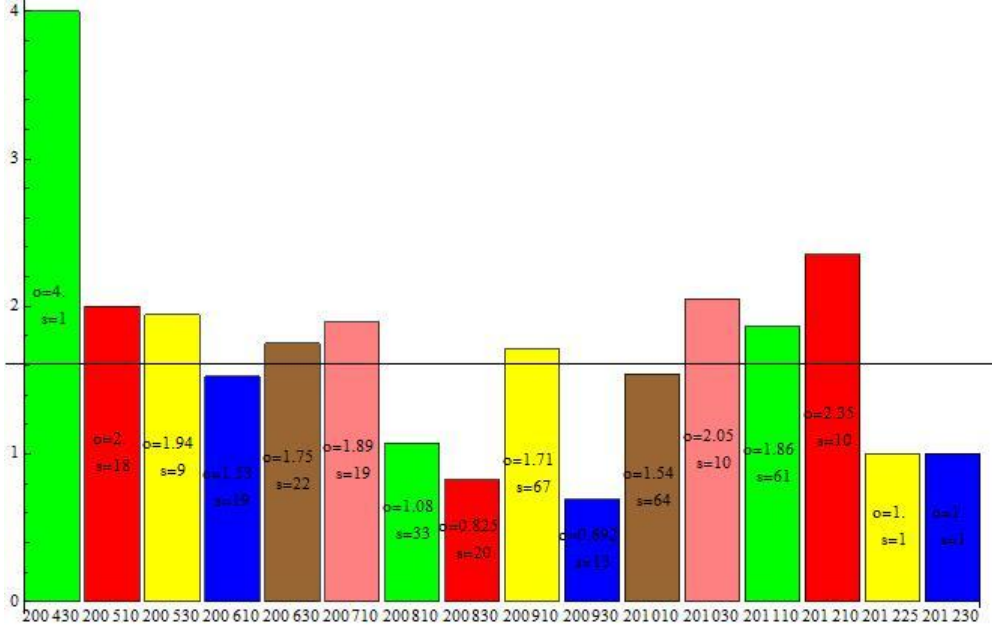
Şekil 4.49: MUH212 dersini ilk kez alanların dönemlere göre ortalaması



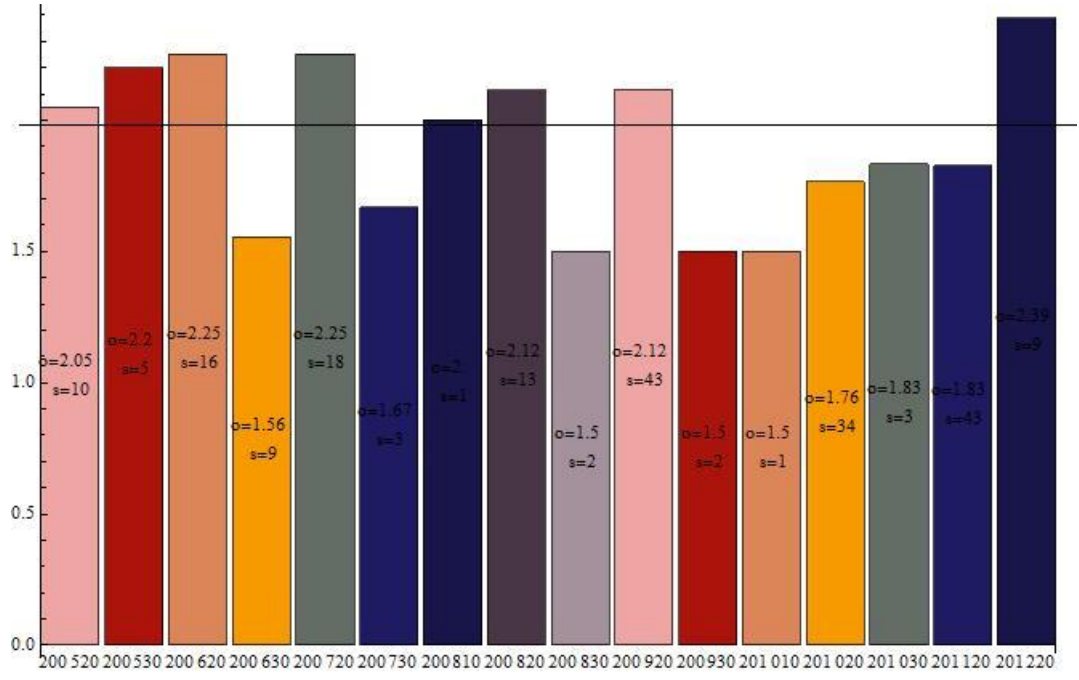
Şekil 4.50: MUH212 dersini alan tüm öğrencilerin dönemlere göre ortalaması



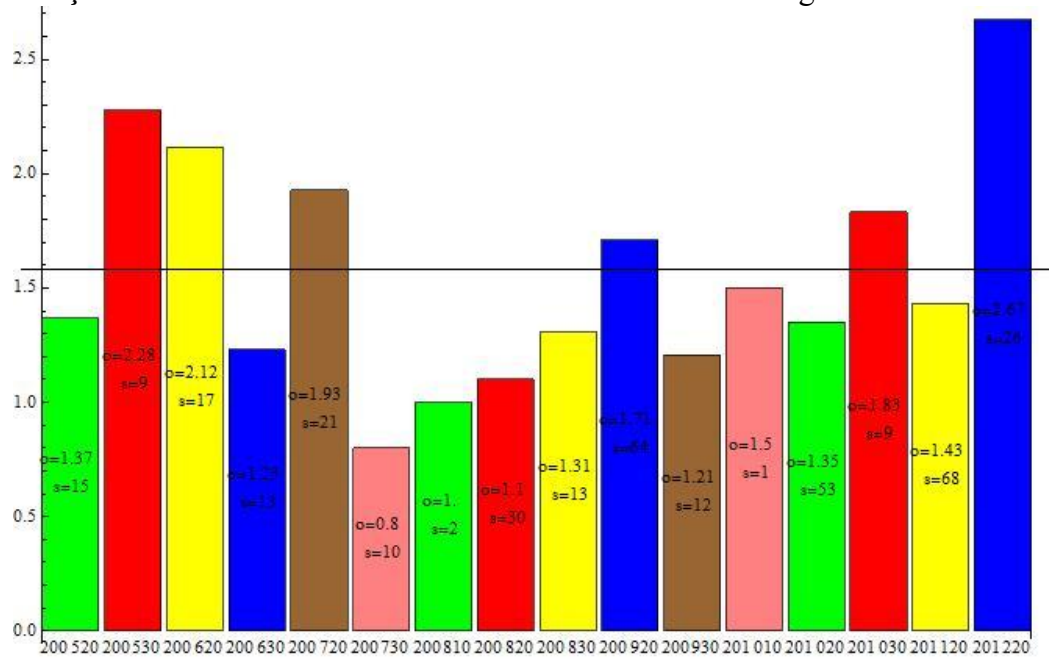
Şekil 4.51: MUH311 dersini ilk kez alanların dönemlere göre ortalaması



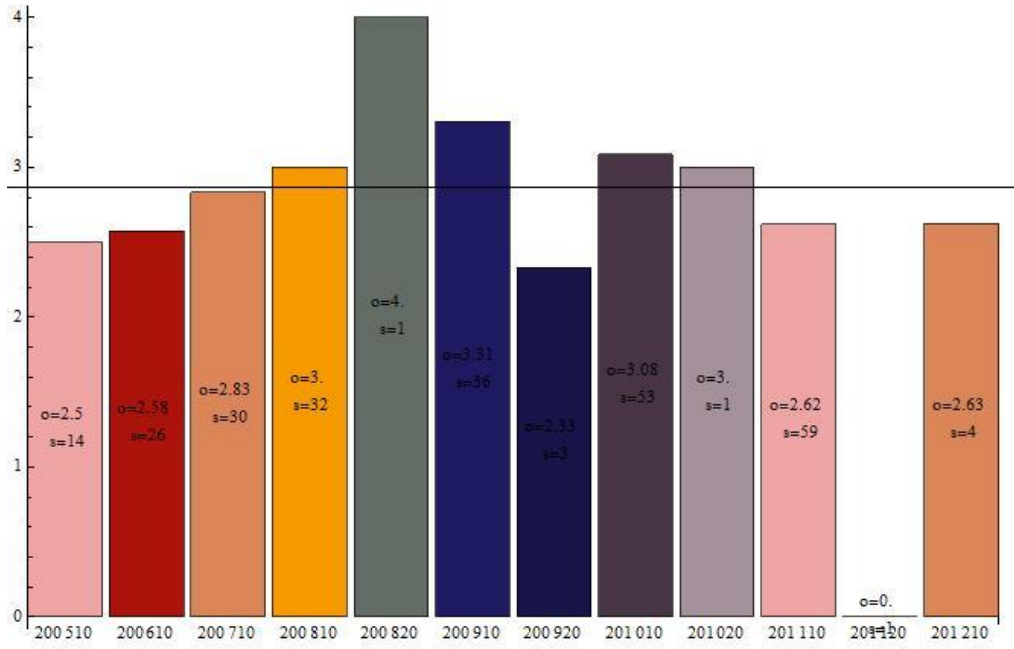
Şekil 4.52: MUH311 dersini alan tüm öğrencilerin dönemlere göre ortalaması



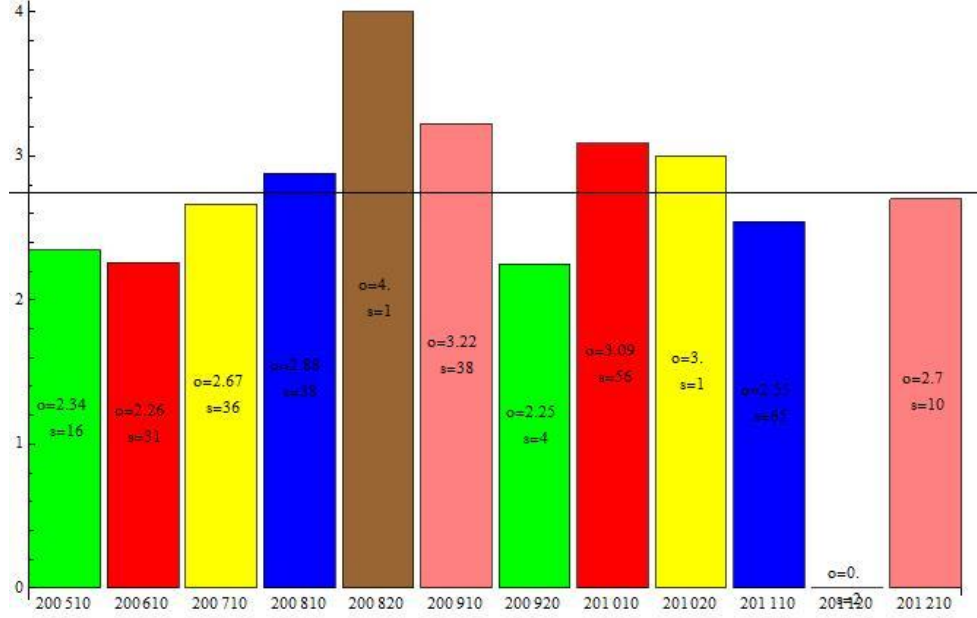
Şekil 4.53: MUH312 dersini ilk kez alanların dönemlere göre ortalaması



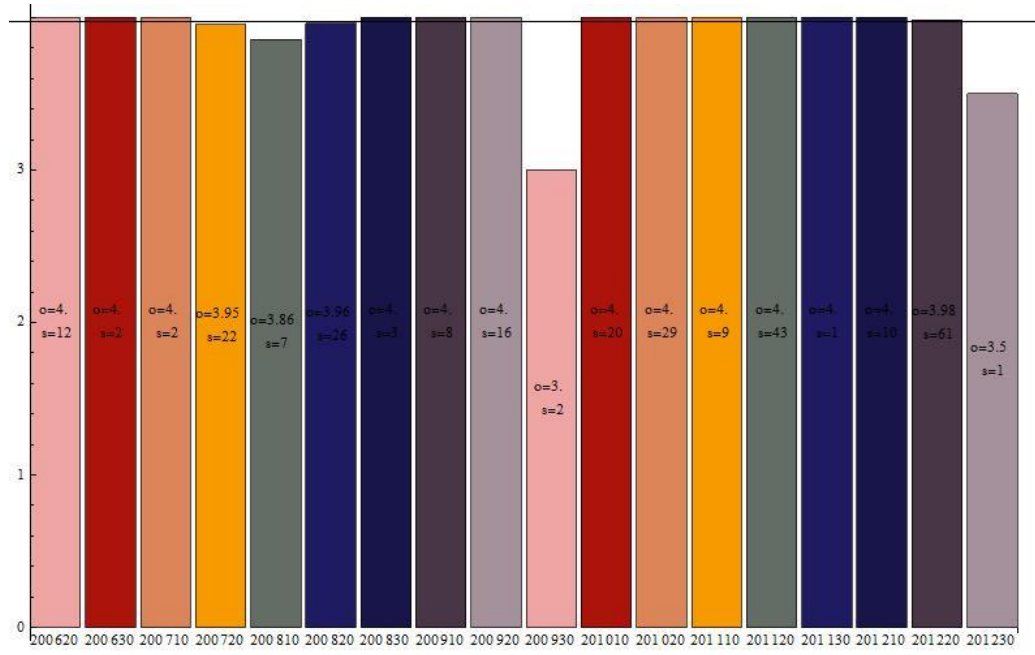
Şekil 4.54: MUH312 dersini alan tüm öğrencilerin dönemlere göre ortalaması



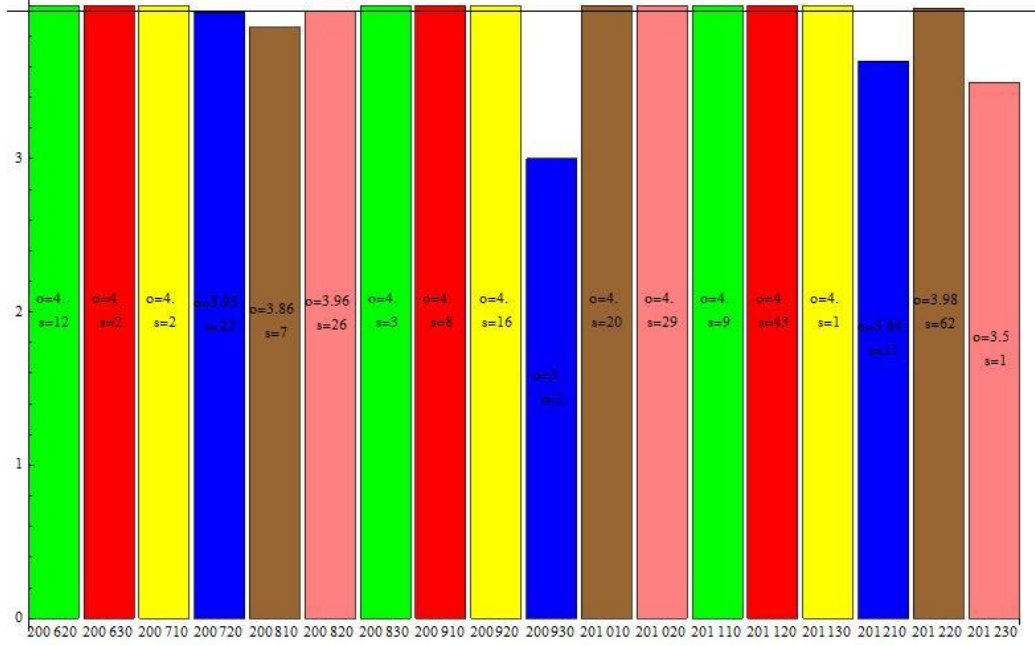
Şekil 4.55: MUH321 dersini ilk kez alanların dönemlere göre ortalaması



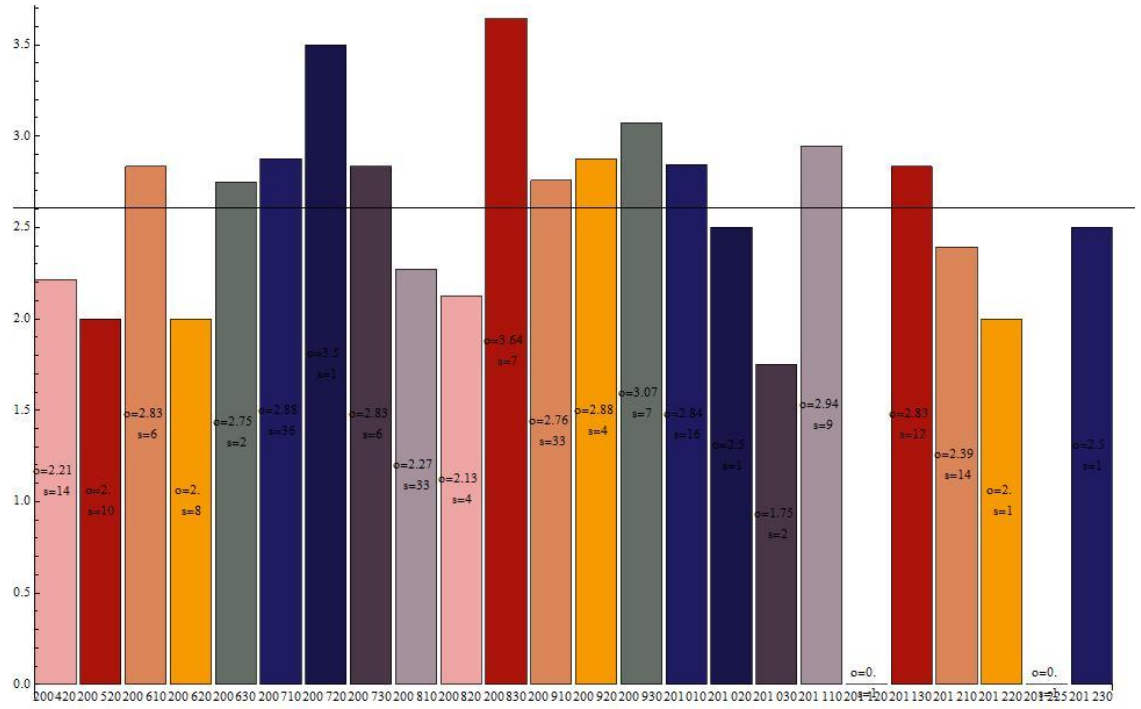
Şekil 4.56: MUH321 dersini alan tüm öğrencilerin dönemlere göre ortalaması



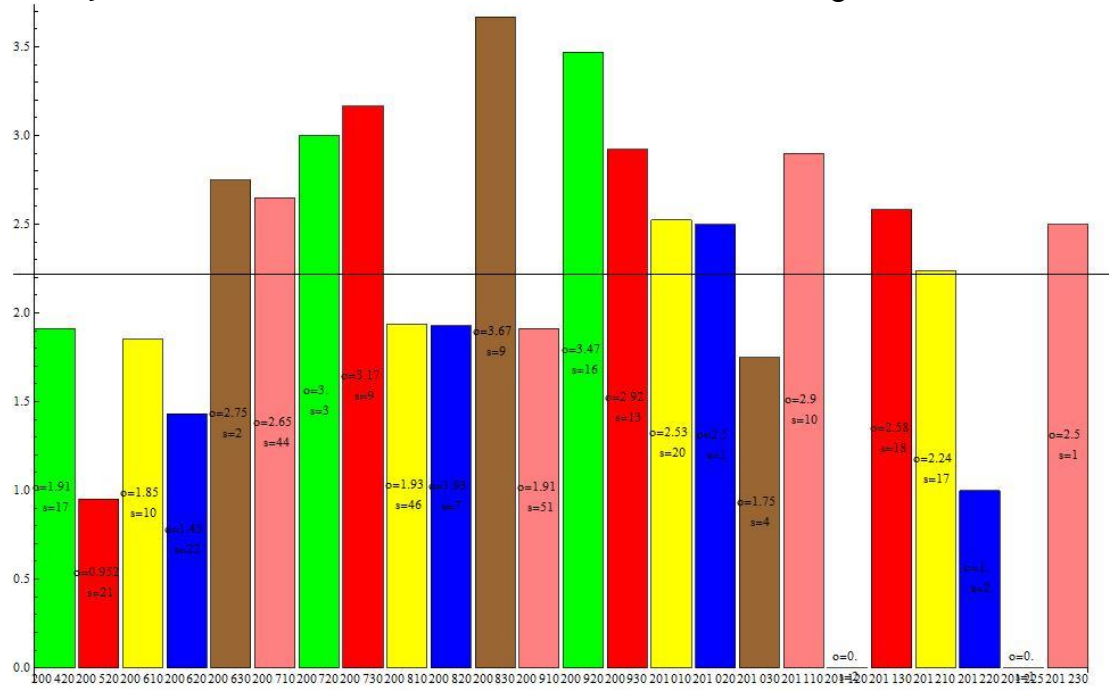
Şekil 4.57: MAT492 dersini ilk kez alanların dönemlere göre ortalaması



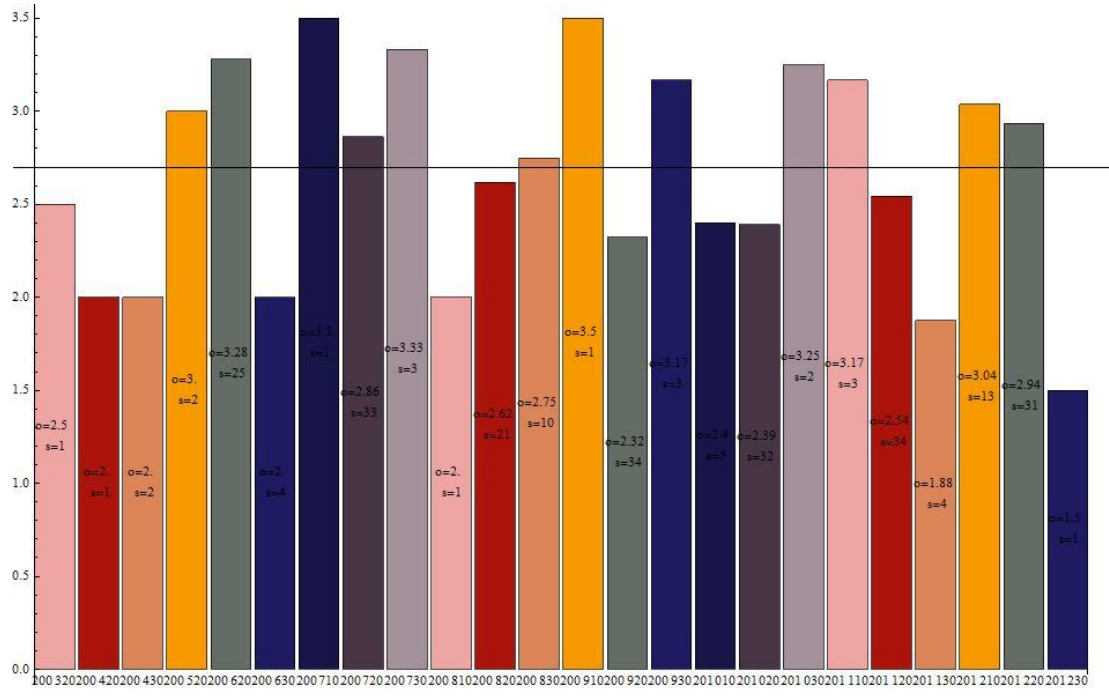
Şekil 4.58: MAT492 dersini alan tüm öğrencilerin dönemlere göre ortalaması



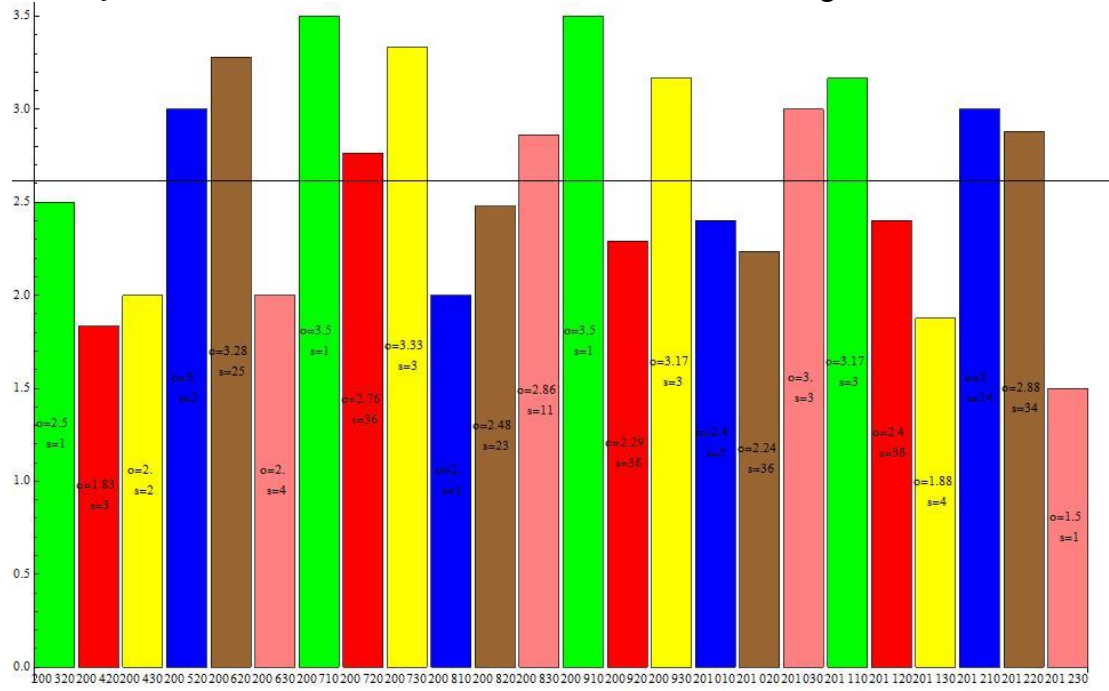
Şekil 4.59: DNK201E dersini ilk kez alanların dönemlere göre ortalaması



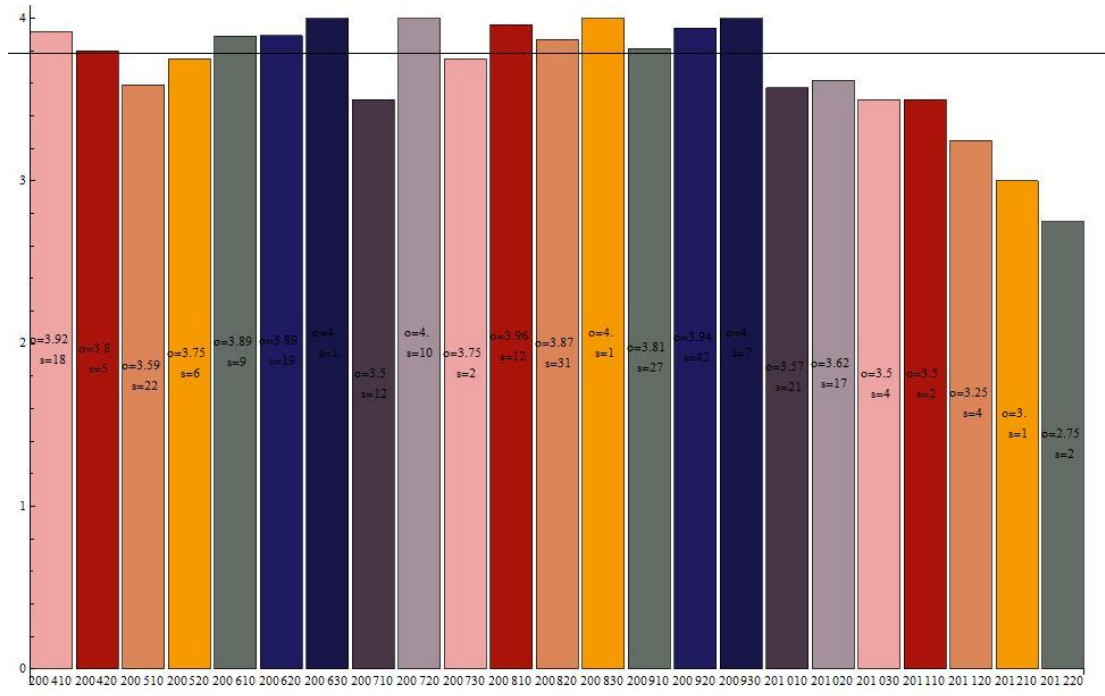
Şekil 4.60: DNK201E dersini alan tüm öğrencilerin dönemlere göre ortalaması



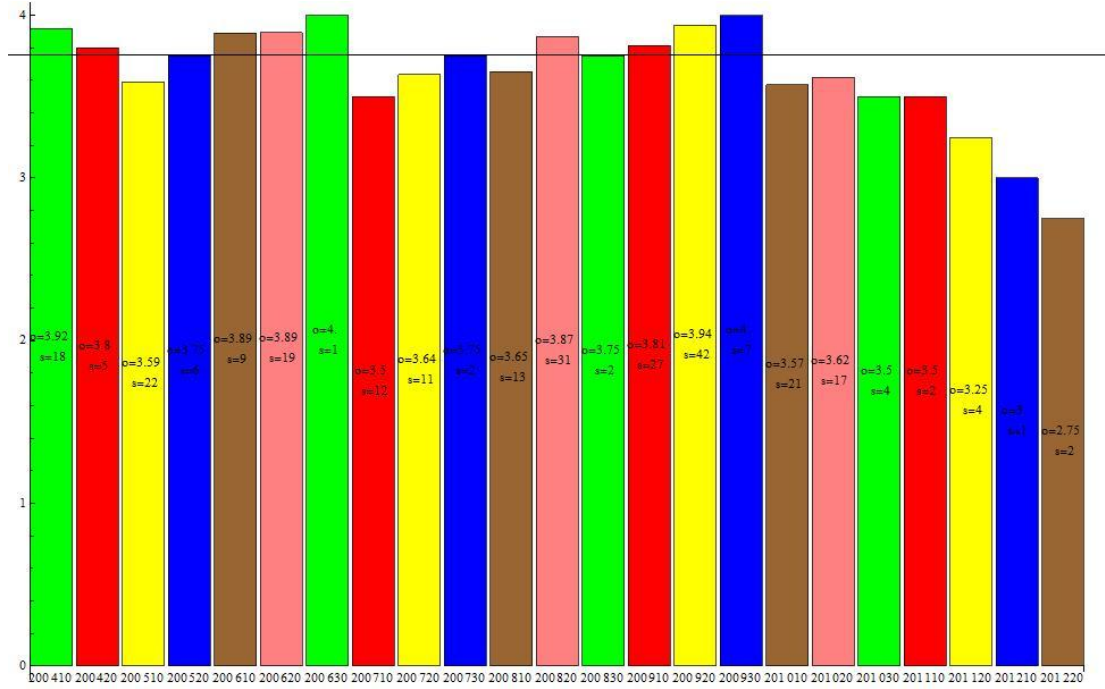
Şekil 4.61: EKO201 dersini ilk kez alanların dönemlere göre ortalaması



Şekil 4.62: EKO201 dersini alan tüm öğrencilerin dönemlere göre ortalaması



Şekil 4.63: ING201 dersini ilk kez alanların dönemlere göre ortalaması

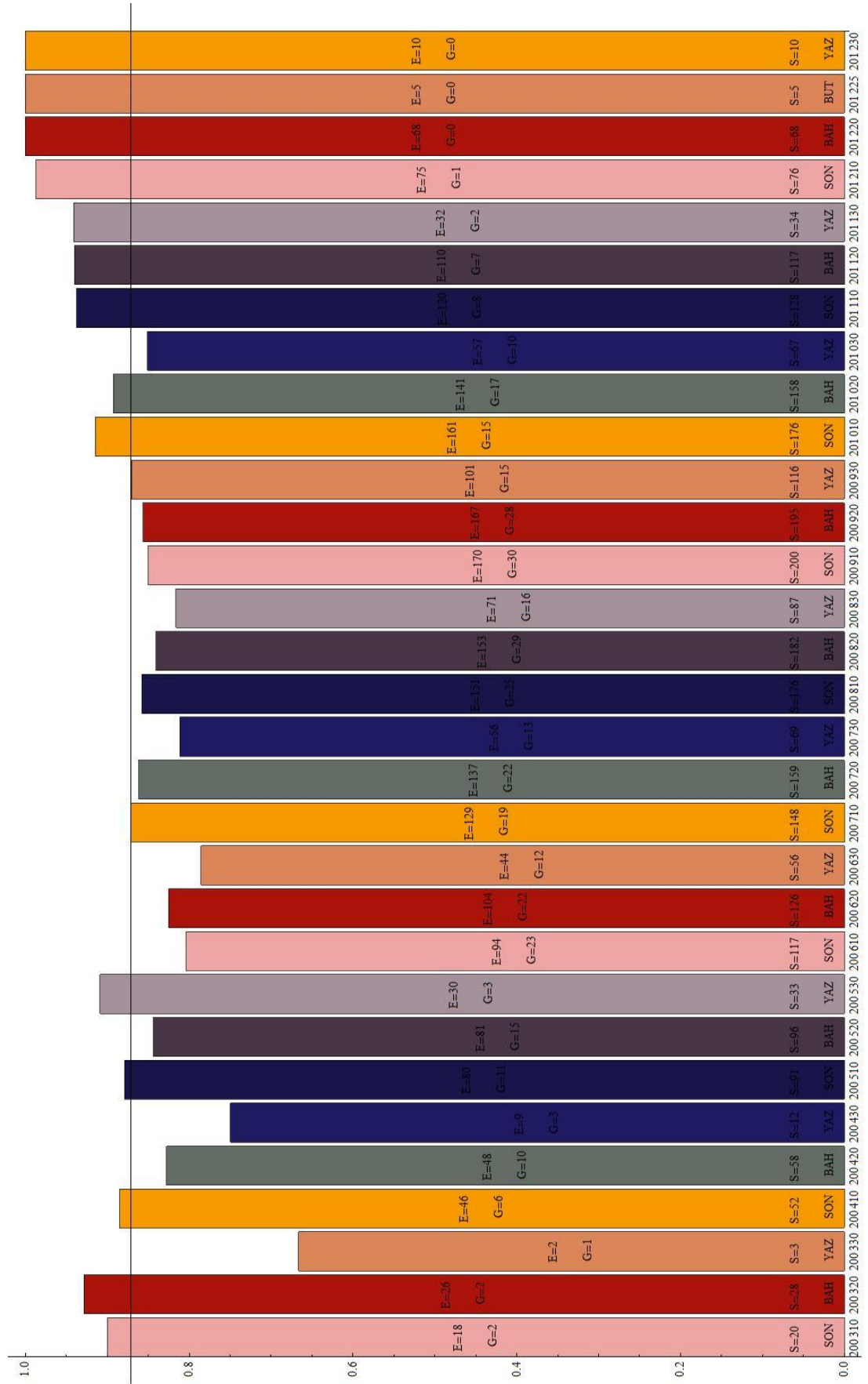


Şekil 4.64: ING201 dersini alan tüm öğrencilerin dönemlere göre ortalaması

Tablo 4.36’da dönemlere göre akademik durum istatistikleri verilmiştir, bu bilgilere göre oluşturulmuş sütun grafiği Şekil 4.65’te verilmiştir. Sütun grafiğinde her bir sütun bir dönemi gösterir. Sütunların üzerinde dönemlerin türü “SON”: Güz Dönemi, “BAH”: Bahar Dönemi, “YAZ”: Yaz Dönemi, “BUT”: Bütünleme Sınavı olarak belirtilmiştir. “S” ile toplam öğrenci sayısı, “E” ile iyi durum öğrenci sayısı, “G” ile gözetim durum öğrenci sayısı gösterilmiştir (Ek-16).

Tablo 4.36: Dönemlere göre akademik durum istatistikleri

	Güz Öğrenci Sayısı			Bahar Öğrenci Sayısı			Yaz Öğrenci Sayısı		
DÖNEM	Toplam	İyi	Gözetim	Toplam	İyi	Gözetim	Toplam	İyi	Gözetim
2003	20	18	2	28	26	2	3	2	1
2004	52	46	6	58	48	10	12	9	3
2005	91	80	11	96	81	15	33	30	3
2006	117	94	23	126	104	22	56	44	12
2007	148	129	19	159	137	22	69	56	13
2008	176	151	25	182	153	29	87	71	16
2009	200	170	30	195	167	28	116	101	15
2010	176	161	15	158	141	17	67	57	10
2011	128	120	8	117	110	7	34	32	2
2012	76	75	1	68	68	0	10	10	0



Şekil 4.65: Dönemlere göre akademik durum sütun grafiği

5. VERİNİN HAZIRLANMASI

İstanbul Teknik Üniversitesi Öğrenci İşleri Daire Başkanlığı'ndan elde edilen Matematik Mühendisliği öğrenci verileri, 2002-2003 Güz Dönemi'nden 2012-2013 Güz Dönemi'ne kadar Matematik Mühendisliği Bölümü'ne kabul edilmiş tüm kabul türlerinde 619 öğrenciden oluşur. Öğrencilerin kabul türlerine göre dağılımı Tablo 5.1'de verilmiştir.

Tablo 5.1: Öğrencilerin kabul türlerine göre dağılımı

Kabul Türü	Öğrenci Sayısı
Standart ÖSS	562
Standart YÖS	12
Ek kontenjan ÖSS	1
Üniversite İçinden Yatay Geçiş	30
Üniversite Dışından Yatay Geçiş	5
5316 No'lu Af Kanunu	1
6111 No'lu Af Kanunu	8

Bu süre içinde lisans eğitime başlamış 574 öğrenci vardır. Öğrencilerin **%7.84'ünün lisans eğitime hiç başlamadığı** görülmektedir. Üniversite içi yatay geçiş türüyle gelen öğrencilerin not dökümlerinde Matematik Mühendisliği Bölümü'ne gelmeden önceki bölümlerde almış oldukları derslerin de bulunduğu, ortalamalarına katıldığı ve dolayısıyla akademik durumlarını etkilediği görülmüştür. Üniversite dışı yatay geçiş öğrencilerinin İTÜ'ye gelmeden önceki üniversitelerinde almış olup İTÜ Matematik Mühendisliği Bölümü dersleri yerine saydırdıkları derslerin M harfi ile işaretlendiği görülmüştür. “5316 No'lu Af Kanunu” ve “6111 No'lu Af Kanunu” kabul türüne sahip öğrencilerin öğretim dönemi sayısı fazlalığı nedeniyle veri setinden silinmesi öngörülmüştür.

Model kurmada kullanılacak öğrenci verileri Mezun Standart ÖSS/YÖS kabul türündeki öğrenciler şeklinde belirlenmiştir.

Mezun Standart ÖSS/YÖS kabul türündeki öğrencilerin transfer bilgilerinde M harf notuna sahip derslerinin olduğu görülmüştür, bu öğrenciler veri setinden çıkarılmıştır.

Yukarıda bahsedilen veri ayıklama işlemlerinden sonra Tablo 5.2’de görüldüğü üzere mezun standart ÖSS/YÖS kabul tipinde 275 öğrenci bulunmaktadır. **Oluşturulan veri madenciliği modelinde mezun standart ÖSS/YÖS kabul tipindeki öğrencilerin bilgileri kullanılmıştır.** Bundan sonraki kısımlarda mezun standart ÖSS/YÖS kabul tipindeki öğrencilere, “Mezun Standart Öğrenci” denilmiştir.

Tablo 5.2: Mezun öğrencilerin kabul türlerine göre dağılımı

Kabul Türü	Mezun Öğrenci Sayısı
Standart ÖSS	270
Standart YÖS	5
Ek kontenjan ÖSS	0
Üniversite İçinden Yatay Geçiş	18
Üniversite Dışından Yatay Geçiş	3
5316 No’lu Af Kanunu	0
6111 No’lu Af Kanunu	0

Mezun standart öğrencilere ait transfer ve tarihçe bilgileri birleştirilmiştir. Bu bilgileri arasında genel not ortalamasına katılmayan T ve E harfleri ile işaretlenmiş dersleri ve JPN, ALM, FRA, BEB gibi kredisiz seçmeli dersleri çıkarılmıştır.

Mezun standart öğrencilerin üniversite öncesi başarılarının değerlendirilebilmesi için ÖSYM bilgisi kullanılmaktadır. İTÜ Öğrenci İşleri’nden elde edilen öğrencilerin ÖSYM verileri, bölüme kabul yılı 2004’den 2012’ye kadar olan öğrenciler olduğundan bölüme 2004 yılı öncesinde giriş yapan öğrenciler için bütünlüğü bozmayacak şekilde başarı sırası bilgisi verilmiş, giriş türü “Genel” olarak belirlenmiştir. Mezun Standart YÖS öğrencileri için ortalama bir ÖSYM başarı sırası değeri üretilmiş, ÖSYM giriş türünde ise “Yabancı Öğrenci” oldukları belirtilmiştir.

Mezun standart öğrencilerin “Regular” olma durumları “Regular” ise “1”, “Irregular” ise “0” şeklinde belirlenmiştir (EK-17, EK-18).

Mezun standart öğrencilerin tüm transfer ve tarihçe bilgileri bir araya getirilerek öğrencilerin öğrenim hayatı boyunca aldığı tüm dersler bir araya getirilmiştir.

Bölüm 1.3’te belirtildiği üzere farklı ders planlarına göre ders alan öğrencilerin not döküm verilerinde bütünlüğü sağlamak amacıyla bir takım düzenlemeler yapılması gerekmiştir. Bu düzenlemeler için Ders Planı 3 baz alınmıştır. MAT141, DNK201, MAT214, MAT242 derslerinin ilgili tablolarda ders planlarında farklı kredilerde olduğu görülmektedir. Yöntemlerin uygulanabilmesi için hazırlanan tablolarda tüm öğrenciler için bu derslerin Ders Planı 3’teki kredileri sayılmıştır.

Tablo 1.2 ve Tablo 1.3’te yer alan “English Course 1” ve “English Course 2” dersleri ING101, ING102, ING103 opsiyonlarına sahiptir. Bir öğrenci bu dersleri sırasıyla ING101, ING102 veya ING102, ING103 şeklinde alabilir. ING101 veya ING103’ten biri zorunlu olarak alındığından “English Course 1” zorunlu dersi “ING101/103” ile isimlendirilmiştir. ING102, tüm öğrenciler tarafından alınmıştır bu sebeple “English Course 2” zorunlu dersi “ING102” ile isimlendirilmiştir.

Tablo 1.3’te görüldüğü gibi Ders Planı 1 ve Ders Planı 2’de MAT142 (Lineer Cebir) dersi yerine MAT122 (Analitik Geometri) dersi bulunmaktadır. Not dökümünde MAT122 olan öğrencilerin bu ders ile ilgili bilgileri, MAT142 dersi yerine sayılmıştır.

Tablo 1.3’ten Ders Planı 1, Ders Planı 2, Ders Planı 3’te “Intro to Sci & Eng Comp” olarak geçen BIL104E, BIL106E ve BIL108E opsiyonlarına sahip ders yerine Ders Planı 4’te BIL104E zorunlu dersinin getirildiği görülmektedir. Dolayısıyla BIL106E ve BIL108E dersleri, BIL104E olarak sayılmaktadır.

MAT322E – Algebra dersinin Tablo 1.4 ve Tablo 1.7’den sadece Ders Planı 1’de 3. sınıf dersi olduğu, diğer ders planlarında 2. sınıf dersi olduğu görülmektedir. Bu sebeple Ders Planı 1’den sorumlu öğrenciler için MAT322E dersine ait dönem bilgisi üzerinde düzenleme yapılmıştır. Bu düzenleme dönem yılının "bir" azaltılması şeklindedir (EK-5).

Tablo 1.4’ten sadece Ders Planı 1’de MUK201E olduğu görülmektedir. MUK201E dersini alan öğrenciler için bu ders ile ilgili bilgiler fazladan alınan seçmeli ders gibi değerlendirilmiştir.

Mezun standart öğrencilerin ders bilgileri değerlendirilirken, öğrencilerin ders planında Türkçe görünen dersleri İngilizce olarak da alabildikleri, İngilizce görünen dersleri ise Türkçe olarak da alabildikleri görülür. Bu tip durumlarda dil ayrımı yapılmaksızın aynı ders olarak işleme devam edilmiştir.

Öğrencilerin not dökümlerinde görülen ders planındaki derslerle denkliği olan dersler, karşılık bulduğu ders gibi sayılmıştır. Örneğin DNK203, DNK201E sayılmıştır. Ayrıca FIZ102'nin sadece 330 numaralı öğrenci tarafından alındığı görülmüştür. Bu öğrencinin FIZ102 dersi için bilgilerinin transfer dosyasından geldiği görüldüğünden FIZ102 dersi FIZ106 yerine sayılmıştır.

BIL101E dersinin başarı değerlendirmesinin zaman içinde değiştirildiği, not karşılığı olmayan BL/BZ harfi yerine harf notu sisteminin getirildiği görülmüştür. Bu sebeple uygulanacak yöntemler için tablo oluşturulurken, BIL101E dersi ile ilgili bu durum göz önünde bulundurulmaktadır.

Uygulanacak yöntemlerde oluşturulan tablolarda, öğrencilerin bulundukları sınıflarda almadıkları zorunlu sınıf derslerinin harf notu bilgilerine “YYY: Yok” ve ders planında bulunan bir dersi öğrenim hayatı boyunca hiç almadıysa “NNN: Hiç Almadı” işaretleri konulmaktadır. İstisnai olarak 60 numaralı öğrencinin ders planında yer almasına rağmen not dökümünde yer almayan MAT142 dersinin harf notu bilgisi de “NNN” ile işaretlenmiştir. Bu öğrencinin bu duruma rağmen nasıl mezun edildiği bilinmemektedir.

Tablo 1.7'den görüldüğü gibi MAT355 – Topoloji Dersi Ders Planı 1'de zorunlu dersler içerisinde yoktur fakat seçmeli ders olarak alınabilmektedir. Ders Planı 1'den sorumlu olup bu dersi seçmeli olarak almayan öğrencilerin de olduğu tespit edilmiştir. Bu öğrenciler için MAT355 ders bilgilerine harf notu yerine, “NNN: Hiç Almadı” işareti verilir (EK-6).

Öğrencilerin transfer bilgilerinden gelen ve ders planında yer almayan seçmeli dersleri ile yüksek lisans ders planına ait dersleri seçmeli ders olarak sayılmaktadır.

Yukarıda ayrıntılı olarak anlatılan düzenlemeler, ekte verilen programlar yardımıyla uygulanarak "mezun standart öğrenci veri seti" hedeflenen veri madenciliği problemini uygulamaya hazır hale getirilmiştir.

6. UYGULAMA VE SONUÇLAR

Bu bölümde Naive Bayes ve birliktelik kuralları algoritmalarının Oracle Veri Madenciliği ile uygulanışı anlatılacaktır. Tez kapsamındaki problem, Matematik Mühendisliği öğrencilerinin derslerden aldıkları harf notları, akademik başarı durumları ile mezuniyet başarı durumlarının tahmin edilmesi ve öğrencilerin bölüm dersleri başarı durumları arasındaki ilişkilerin varsa ortaya çıkarılmasıdır. Benzer çalışma Khair, Zakhem ve Moucary tarafından 2012 yılında yayınlanmıştır [7]. Bu tez çalışmasında [7, 8] baz alınarak öğrencilerin derslerden aldıkları harf notları, akademik başarı durumları ile mezuniyet başarı durumlarının tahmin edilmesi problemi için iki farklı şekilde tablo oluşturulmuştur ve bu tablolara Naive Bayes algoritması uygulanmıştır. Öğrencilerin bölüm dersleri arasındaki ilişkilerin çıkarılması problemi için ise birliktelik kuralları Apriori algoritması uygulanmıştır.

Burada öncelikli olarak teze başlarken konuyu öğrenme sürecinde yapılan hatalı bir uygulama sonucu verilmiştir. Sonrasında ise sunulan problemin çözümü amacıyla doğru modelin oluşturulma, uygulanma ve test süreçleri Naive Bayes ve Birliktelik Kuralları algoritmaları için anlatılmıştır. Son olarak kullanılan modelin test sürecinde ki başarısı göz önünde tutularak, elde edilen sonuçların yorumlanması ve bunların nasıl değerlendirilebileceği üzerinde durulmuştur.

6.1. Hatalı bir Naive Bayes Uygulaması

Yukarıda verilen problemin çözümü doğrultusunda, ilk olarak birinci sınıf verileri kullanılarak mezun olup olmama ve mezuniyet başarısının tahmini için bir model oluşturulmuştur. Bu modelin uygulanması sonucunda yapılan tahminlerin çok yüksek çıktığı görülmüştür. Fakat bu sonuçların yüksek çıkması, modelde kullanılan tablonun oluşturulma mantığının hatalı olmasından dolayıdır. Bu sebeple ikinci ve üçüncü sınıf için aynı uygulamaya devam edilmemiştir.

Öncelikle öğrencilerin ders planında yer alan birinci sınıf dersleri için son başarı notu kullanıldığından bu derslerden **herhangi birini en son aldığı döneme kadar**

öğrenci birinci sınıf sayılmıştır. Ancak bu durumda örneğin birinci sınıf dersi olan MAT111 dersini mezuniyet döneminden hemen önce tekrar alan bir öğrencinin tüm bilgisi mezuniyet başarısı tahmini için kullanılmaktadır. Dolayısıyla sadece birinci sınıf bilgilerini kullanarak tahminde bulunma amacının dışına çıkılmakta hatta öğrencinin gerçekte birinci sınıf olduğu dönemde var olmayan bir bilgi modelde kullanılmış olmakta ve bu da hataya neden olmaktadır.

Bu **hatalı** model uygulandığında tahminin doğruluğunun Şekil 6.1’de de görüldüğü gibi %97,8182 olduğu görülmüştür.

Performance

Performance Matrix

Lift

Profit

Average Accuracy: 97,9131

Overall Accuracy: 97,8182

Target Value	Total Case Count	Correct Predictions %	Cost	Cost %
1	69	98,5507		
2	117	97,4359		
3	89	97,7528		

Şekil 6.1: Hatalı uygulamanın güvenilirliği

Şekil 6.2’den de görüldüğü üzere yalnız 6 öğrencinin mezuniyet başarıları yanlış tahmin edilmiştir.

Performance Matrix: Rows=Actual, Columns=Predicted: ☒ Show totals and cost:						
	1	2	3	Total	Correct %	Cost
1	68	1	0	69	98,5507	
2	3	114	0	117	97,4359	
3	0	2	87	89	97,7528	
Total	71	117	87	275		
Correct %	95,7746	97,4359	100			
Cost						

Şekil 6.2: Hatalı uygulamanın “Performance Matrix” sonucu

Kurulan model, Şekil 6.1’de de görüldüğü üzere %97,8 gibi çok yüksek bir doğrulukta çıkmıştır. Bahsedildiği gibi öğrencilerin birinci sınıf sayıldıkları dönemin ucu açık bırakılması, öğrencilerin mezuniyet dönemlerine çok yakın zamandaki bilgilerinin de kullanılmasına neden olmuştur bu da modelin yüksek doğrulukta elde edilmesine sebep olmuştur. Birinci sınıf için gerçekte var olmayan daha sonraki bilgilerin kullanılması modelin gerçekte uygulanabilirliği ortadan kaldırmaktadır.

6.2. Sınıflara Göre Mezuniyet Başarısının Tahmini

Bu bölümde İstanbul Teknik Üniversitesi Öğrenci İşleri Daire Başkanlığı'ndan elde edilen Matematik Mühendisliği mezun öğrenci verileri kullanılarak, mezuniyet başarı durumları üzerine yapılan modelleme anlatılmaktadır. Giriş bölümünde de değinildiği üzere kurulan modelde öğrencilerin mezuniyet başarı durumlarının sınıf bilgileriyle tahmin edilmesi amaçlanmıştır.

Öğrencilerin tahmin edilecek mezuniyet başarı durumları, öğretimlerini yaz okulları ile beraber 12 yarıyıl sonunda mezun olamayan öğrenciler, öğretimlerini yaz okulları ile beraber en fazla 12 yarıyıl da bitirmiş “düşük”, “orta”, “yüksek” başarı sınıflarında mezun olmuş öğrenciler şeklinde belirlenmiştir.

Kurulan model Bölüm 6.1’de anlatıldığı gibi, öğrenci bilgilerinin sınıflara göre değerlendirilmesine dayanmaktadır. “Regular” öğrenciler, öğrenimlerine başladığı güz dönemini takip eden bahar dönemi ve (eğer ders aldıysa) yaz okulu; “Irregular” öğrenciler ise öğrenimlerine başladığı bahar dönemini takip eden (eğer ders aldıysa) yaz okulu ve güz dönemi boyunca birinci sınıf sayılmaktadırlar. Devamındaki 3 dönem boyunca ikinci sınıf ve aynı şekilde devamındaki 3 dönem boyunca üçüncü sınıf sayılmaktadırlar.

Öğrencilerin bulundukları sınıf boyunca bu sınıflara ait derslerinden aldıkları ilk ve son harf notları, her bir dersi kaç kere tekrar ettiği ve akademik durumlarının kaç kere iyi durum ve/veya gözetim durumu olduğu bilgileri hedef sütunun tahmin edilmesinde kullanılmıştır. Öğrencinin bulunduğu sınıfta tekrar alıp notunu yükselttiği/düşürdüğü dersler, bulunduğu sınıfta ilk kez aldığı sınıf dersleri dışında alınan ortalamasını yükseltecek/düşürecek etkiye sahip ders bilgileri belirlenmiş ve tahmin edici bilgi olarak kullanılmıştır. Bunların yanı sıra öğrencinin “Regular” olma durumu, ÖSYM başarı sırası ve ÖSYM giriş türü bilgileri de mezuniyet başarısını tahmin etmede kullanılmıştır.

6.2.1 Birinci sınıf verilerine göre mezuniyet başarısının tahmini

Öğrencilerin birinci sınıf bilgileriyle mezuniyet başarılarının tahmin edilmesi için kurulan modelde kullanılan veri tablosunun sütun açıklamaları Tablo 6.1’de verilmiştir.

Tablo 6.1: Birinci sınıf tablosunun sütun açıklamaları

SÜTUN NO	SÜTUN ADI	SÜTUN AÇIKLAMASI
1	OGRENCI_NO	Öğrencinin ID'sini belirtir. "8" ile "497" arasında değer alır.
2	MAT111_TEKRAR_SAYISI	MAT111 dersini tekrar etme sayısını belirtir.
3	MAT111_ILK_HARF_NOTU	MAT111 dersinden ilk aldığı harf notunu belirtir. "AA", "BA", "BB", "CB", "CC", "DC", "DD", "FF", "VF", "YYY" değerlerini alır.
4	MAT111_SON_HARF_NOTU	MAT111 dersinden son aldığı harf notunu belirtir. "AA", "BA", "BB", "CB", "CC", "DC", "DD", "FF", "VF", "YYY" değerlerini alır.
5-46	Bu sütunlar aşağıdaki dersler için sırasıyla Sütun 2, Sütun 3 ve Sütun 4'teki gibi adlandırılmıştır: MAT112, MAT141, MAT142, MAT175, FIZ101E, FIZ101EL, FIZ106, FIZ106L, KIM101E, KIM101EL, BIL101E, BIL104E, ING101, ING102	Sütunlar için sırasıyla Sütun 2, Sütun 3, Sütun 4'teki Sütun Açıklamaları geçerlidir. BIL101E_ILK_HARF_NOTU ve BIL101E_SON_HARF_NOTU sütunları için Sütun 2 ve Sütun 3 Sütun Açıklamaları'nda verilen değerlere ek olarak "BL", "BZ" değerleri dahildir.
47	TEKRAR_ARTIM_SAYI	Birinci sınıf boyunca tekrar ettiği derslerden notunu yükselttiklerinin sayısını gösterir. Tüm dersler için dersin tekrar sayısı ile dersin kredisinin çarpılmasından elde edilen değerler toplamıdır.
48	TEKRAR_ARTIM_NOT	Birinci sınıf boyunca tekrar ettiği derslerden notunu yükselttiklerinin getirdiği puanı gösterir. Tüm dersler için dersin yükselme notunun dersin kredisi ile çarpılması ile elde edilen değer toplamıdır.
49	TEKRAR_AZALIM_SAYI	Birinci sınıf boyunca tekrar ettiği derslerden notunu düşürdüklerinin sayısını gösterir. Tüm dersler için dersin tekrar sayısı ile dersin kredisinin çarpılmasından elde edilen değerler toplamıdır.
50	TEKRAR_AZALIM_NOT	Birinci sınıf boyunca tekrar ettiği derslerden notunu düşürdüklerinin getirdiği puanı gösterir. Tüm dersler için dersin yükselme notunun dersin kredisi ile çarpılması ile elde edilen değer toplamıdır.
51	IYI_DURUM	Birinci sınıf boyunca akademik durumunun "iyi durum" olduğu dönem sayısını belirtir.
52	GOZETIM_DURUM	Birinci sınıf boyunca akademik durumunun "gözetim durumu" olduğu dönem sayısını belirtir.
53	GOZETIM_PUANI	Akademik durumu belirten puandır. IYI_DURUM sütun bilgisinden GOZETIM_DURUM sütun bilgisinin çıkarılması ile bulunur. "-3" ile "3" arasında değer alır.
54	REGULAR_OLMA	Regular veya Irregular Olma durumunu belirtir.

		“Regular”:1, “Irregular”:0 değerlerini alır.
55	OSYM_BASARI_SIRASI	ÖSYM tarafından yapılan sınav sonucuna göre başarı sırasını belirtir.
56	OSYM_GIRIS_TURU	ÖSYM tarafından yapılan sınav sonucuna göre okula kabul ediliş durumunu belirtir.
57	MEZUNİYET_BASARISI	Mezuniyet başarı durumunu belirtir. “0”:mezun değil, “1”:düşük başarılı mezun, “2”:orta başarılı mezun, “3”:yüksek başarılı mezun değerlerini alır.

6.2.1.1 Veri tablosunun hazırlanması

Bölüm 5’te anlatılan düzenlemelerin uygulandığı veriler üzerinde Ekler bölümündeki fonksiyonlar birinci sınıf için çalıştırılarak Tablo 6.1’de görülen sütunlar oluşturulmuştur.

Tablo 6.1’de Sütun 2 ile Sütun 46 arasında görüldüğü gibi birinci sınıf dersleri ile ilgili dersten alınan ilk harf notu bilgilerine, son harf notu bilgilerine ve dersin tekrar sayısı bilgileri elde edilmiştir. Sütun 47 ile Sütun 50 arasında görüldüğü üzere birinci sınıf dönemlerinde tekrar edilen dersler ile ilgili bilgiler elde edilmiştir. Tekrar edilen dersler ile ilgili bilgilere, BIL101E’nin not karşılığı olmayan BL/BZ harfleriyle notlandırılması nedeniyle BIL101E ile ilgili tekrar bilgisi dahil edilmemiştir. Tablo 6.1 Sütun 51, Sütun 52 ve Sütun 53’te görülen akademik durum bilgileri elde edilmiştir. Öğrencilerin birinci sınıfta birinci sınıf dersleri dışında ilk kez aldıkları ders bilgileri, ihmal edilmiştir. Tablo 6.1’de Sütun 54, Sütun 55 ve Sütun 56’da görülen “Regular” olma, ÖSYM başarı sırası, ÖSYM giriş türü bilgileri elde edilmiştir.

Mezuniyet başarısı Tablo 6.1 Sütun 57’de açıklandığı gibi “0”: mezun değil, “1”: düşük başarılı mezun, “2”: orta başarılı mezun, “3”: yüksek başarılı mezun değerlerini alır. Bu sınıflar ve sınırları "K-Means" algoritması ile mezuniyet ortalaması 2.00 ile 2.67 arasında olan “düşük başarılı mezun”, 2.67 ile 3.15 arasında olan “orta başarılı mezun”, 3.15 üzeri olan “yüksek başarılı mezun” olarak belirlenmiştir.

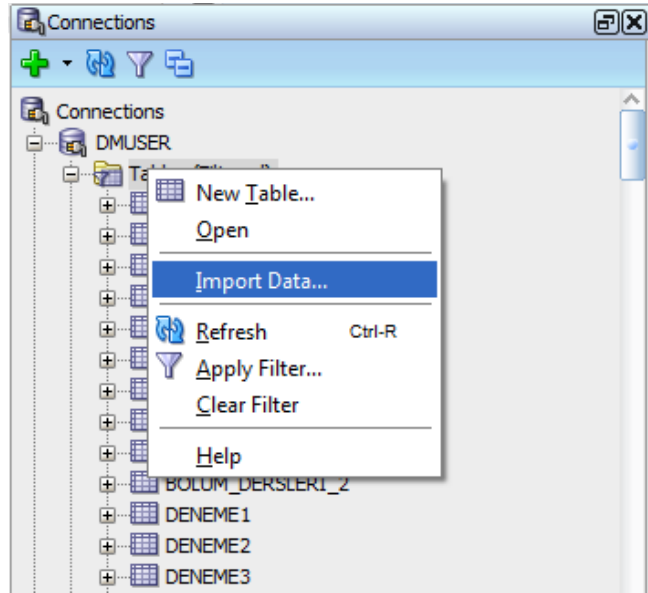
6.2.1.2 Model için tablo oluşturma

Birinci sınıf verilerine göre mezuniyet başarısının tahmini için oluşturulacak modelde kullanılmak üzere hazırlanan tablo “TABLO1_NB.xlsx” dosyasında bulunmaktadır (Şekil 6.3).

	A	B	C	D	E	F	G	H	I
1	OGRENCI_NO	MAT111_TKRAR_SAYISI	MAT111_ILK_HARF_NOTU	MAT111_SON_HARF_NOTU	MAT112_TKRAR_SAYISI	MAT112_ILK_HARF_NOTU	MAT112_SON_HARF_NOTU	MAT141_TKRAR_SAYISI	MAT141
2	8	1	BA	BA	1	BB	BB	1	BA
3	9	1	BB	BB	0	YYY	YYY	1	BA
4	10	1	BA	BA	1	BB	BB	1	AA
5	11	1	DD	DD	1	DC	DC	1	DD
6	12	1	BB	BB	1	CC	CC	1	CC
7	13	1	DD	DD	1	DC	DC	1	CB
8	14	1	DC	DC	1	CC	CC	1	CB
9	15	1	CB	CB	1	CC	CC	1	CB
10	16	1	DD	DD	1	FF	FF	1	FF
11	17	1	CC	CC	1	CC	CC	1	CC
12	18	1	DD	DD	1	DD	DD	1	DD
13	19	1	FF	FF	1	CC	CC	1	BB
14	20	1	BA	BA	0	YYY	YYY	1	CC
15	21	1	CB	CB	1	CB	CB	1	BB
16	22	1	CC	CC	1	CB	CB	1	CB
17	23	1	DD	DD	0	YYY	YYY	1	BB
18	24	1	DD	DD	1	CC	CC	1	CB
19	25	1	DD	DD	1	DD	DD	1	DC
20	26	1	DC	DC	1	CC	CC	1	CB
21	27	1	CC	CC	1	DC	DC	1	CC
22	28	1	FF	FF	0	YYY	YYY	1	DC
23	29	1	FF	FF	0	YYY	YYY	0	YYY
24	30	1	CB	CB	1	CC	CC	1	CC
25	31	1	CC	CC	1	FF	FF	1	FF

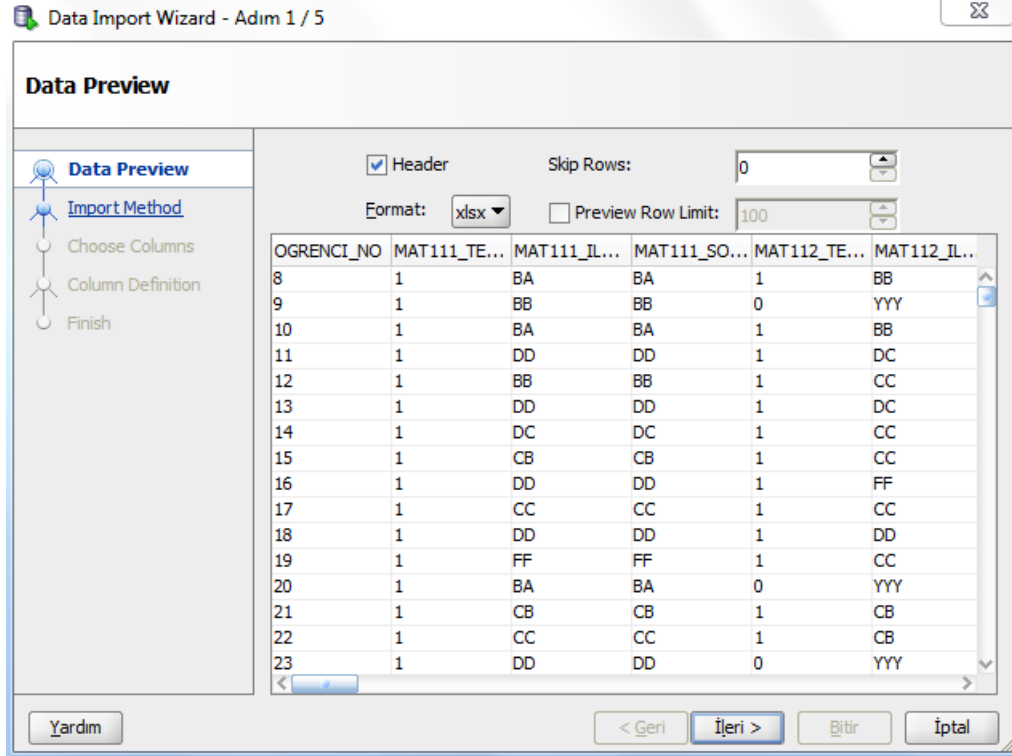
Şekil 6.3: “TABLO1_NB.xlsx” dosyasından bir görüntü

Şekil 6.4’te görüldüğü gibi “DMUSER” altında sağ tıklayarak “Import Data” seçilir [9]. Bu seçenek istenen dosyadan yeni oluşturulacak tabloya veri yüklenmesini sağlamaktadır. ve “BIRINCI_SINIF_NB” şeklinde isimlendirilmiştir.



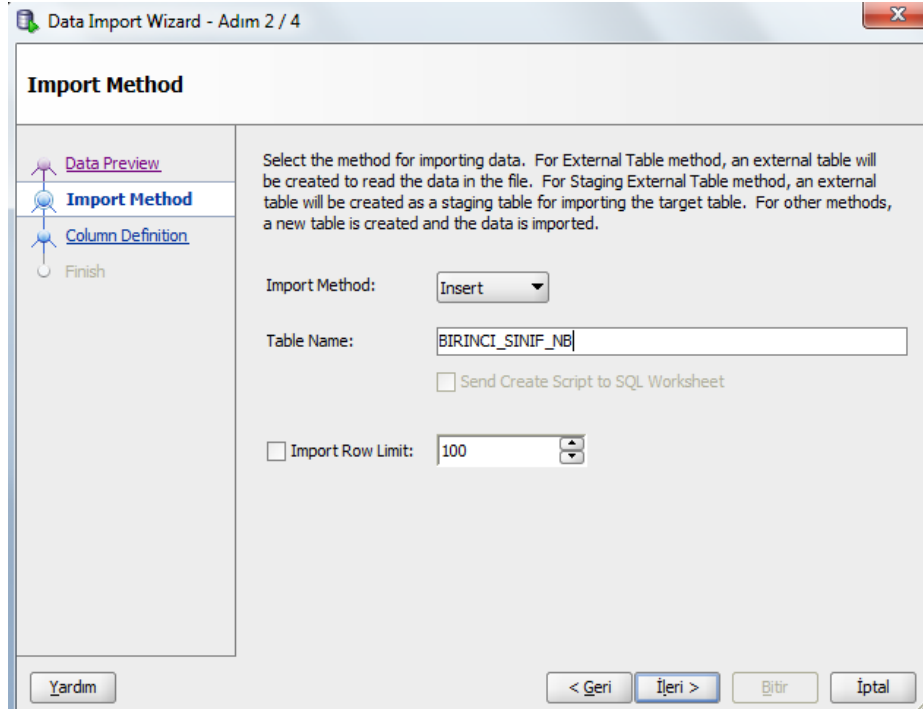
Şekil 6.4: “Import Data” veri aktarma seçeneği

Yükleme yapılacak “TABLO1_NB” dosyası seçtikten sonra “Data Preview” adımımda “Preview Rol Limit” seçeneği verinin tümünü yüklemek için kaldırılır (Şekil 6.5).



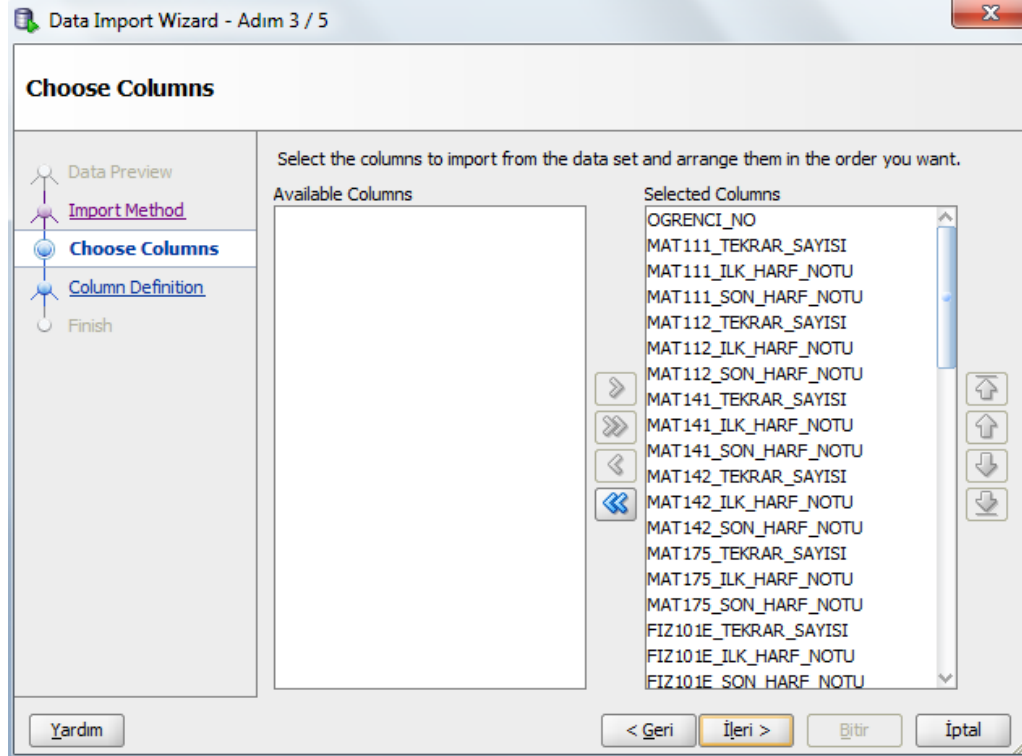
Şekil 6.5: Veri aktarma işlemi verinin görüntülenmesi

“Import Method” adımıyla oluşturulacak tablo isimlendirilir (Şekil 6.6). Birinci sınıf verileri ile oluşturulacak modelde kullanılacak olan bu tabloya “BIRINCI_SINIF_NB” adı verilmiştir.



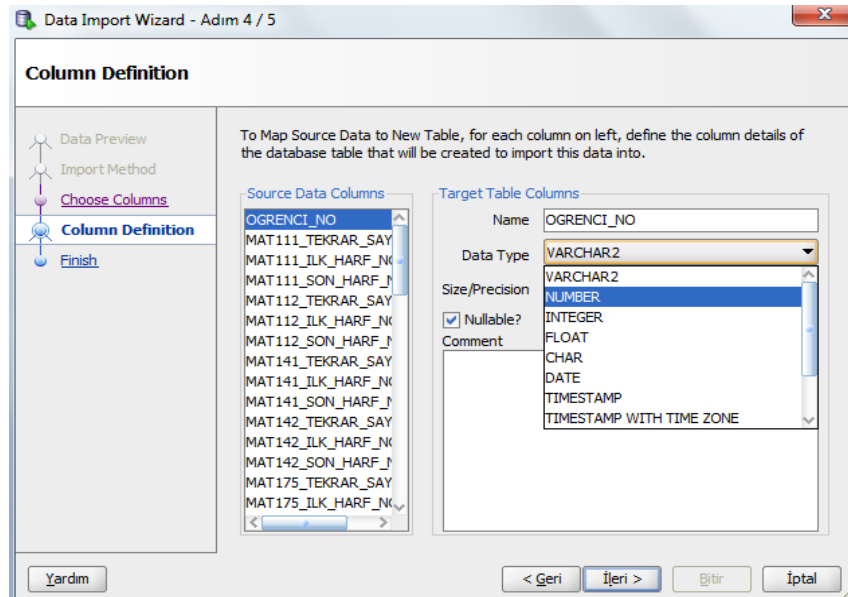
Şekil 6.6: Tablo isimlendirme

Sonraki adım olan “Choose Columns”da tablodan hangi sütunların aktarılacağı seçilir (Şekil 6.7).



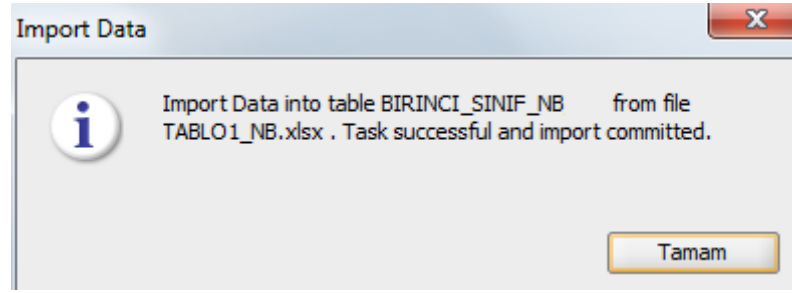
Şekil 6.7: Tahmin edici sütunların belirlenmesi

Kaynak veri tablosunun sütunları ile oluşturulacak tablonun sütunları “Column Definition” adımında eşleştirilir. Veri tipleri her bir sütun için verilerin içeriğine göre düzenlenmiştir (Şekil 6.8).



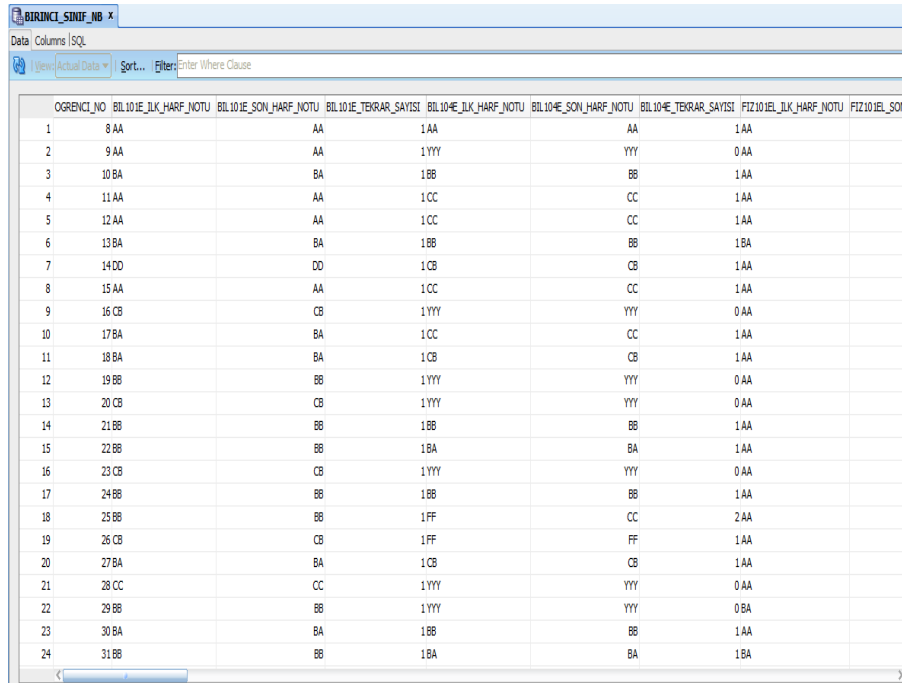
Şekil 6.8: Sütun bilgileri düzenleme

Veri sorunsuz yüklendiğinde Şekil 6.9’daki bilgi kutusu ekrana gelir.



Şekil 6.9: Veri yüklemenin başarılı olduğunu gösteren uyarı

Model kurmak için gerekli veriler yukarıda anlatılan adımlarla Oracle Veri Tabanı’nda “BIRINCI_SINIF_NB” tablosuna aktarılmıştır (Şekil 6.10).



OGRENCI_NO	BIL10 IE_ILK_HARF_NOTU	BIL10 IE_SON_HARF_NOTU	BIL10 IE_TKRAR_SAYISI	BIL10 IE_ILK_HARF_NOTU	BIL10 IE_SON_HARF_NOTU	BIL10 IE_TKRAR_SAYISI	FIZ10 IEL_ILK_HARF_NOTU	FIZ10 IEL_SON_HARF_NOTU
1	8 AA	AA	1 AA	AA	1 AA			
2	9 AA	AA	1 YYY	YYY	0 AA			
3	10 BA	BA	1 BB	BB	1 AA			
4	11 AA	AA	1 CC	CC	1 AA			
5	12 AA	AA	1 CC	CC	1 AA			
6	13 BA	BA	1 BB	BB	1 BA			
7	14 DD	DD	1 CB	CB	1 AA			
8	15 AA	AA	1 CC	CC	1 AA			
9	16 CB	CB	1 YYY	YYY	0 AA			
10	17 BA	BA	1 CC	CC	1 AA			
11	18 BA	BA	1 CB	CB	1 AA			
12	19 BB	BB	1 YYY	YYY	0 AA			
13	20 CB	CB	1 YYY	YYY	0 AA			
14	21 BB	BB	1 BB	BB	1 AA			
15	22 BB	BB	1 BA	BA	1 AA			
16	23 CB	CB	1 YYY	YYY	0 AA			
17	24 BB	BB	1 BB	BB	1 AA			
18	25 BB	BB	1 FF	CC	2 AA			
19	26 CB	CB	1 FF	FF	1 AA			
20	27 BA	BA	1 CB	CB	1 AA			
21	28 CC	CC	1 YYY	YYY	0 AA			
22	29 BB	BB	1 YYY	YYY	0 BA			
23	30 BA	BA	1 BB	BB	1 AA			
24	31 BB	BB	1 BA	BA	1 BA			

Şekil 6.10: “BIRINCI_SINIF_NB” tablosundan bir görüntü

6.2.1.3 Model oluşturma

“BIRINCI_SINIF_NB” tablosu üzerinde Naive Bayes yöntemi uygulanacaktır. Yöntemin oluşturuluşu Bölüm 3.1.3’te anlatıldığı gibidir. İş akış diyagramına eklenen “Data Source” düğümüne “BIRINCI_SINIF_NB” tablosu yüklenir ve bu düğüm eklenen “Classification” düğümüne bağlanır. “Classification” düğümünde yapılan düzenlemelerde, yöntem Naive Bayes ve “Target” yani hedef sütun “MEZUNİYET_BASARISI”, “CaseID” ise “OGRENCI_NO” olarak seçilmiştir (Şekil 6.11).

Target: MEZUNİYET_BASARISI

Case ID: OGRENCI_NO

Name	Algorithm	Date	Auto
CLAS_NB_1_45	Naive Bayes	06.05.2013 00:54	☐

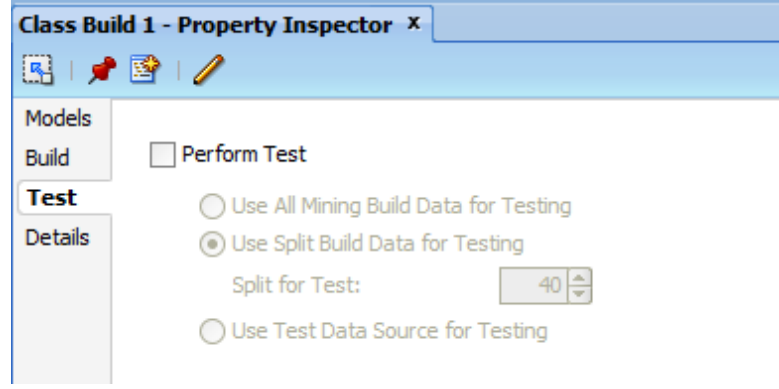
Şekil 6.11: Model oluşturmada ayarlar

“Advanced Settings” (Gelişmiş ayarlar) bölümünde “MEZUNİYET_BASARISI” ve “OGRENCI_NO” sütunları hariç tüm sütunların girdi olarak verilmesine ve bu sütunlar için “Auto Prep” kısmının seçili olmamasına dikkat edilmiştir (Şekil 6.12).

Attributes	Data Type	Input	Mining Type	Auto Prep
☒ KIM101E_SON_H...	VARCHAR2	☒	Categorical	☐
☒ KIM101E_TEKRA...	NUMBER	☒	Numerical	☐
☒ MAT111_ILK_HA...	VARCHAR2	☒	Categorical	☐
☒ MAT111_SON_H...	VARCHAR2	☒	Categorical	☐
☒ MAT111_TEKRAR...	NUMBER	☒	Numerical	☐
☒ MAT112_ILK_HA...	VARCHAR2	☒	Categorical	☐
☒ MAT112_SON_H...	VARCHAR2	☒	Categorical	☐
☒ MAT112_TEKRAR...	NUMBER	☒	Numerical	☐
☒ MAT141_ILK_HA...	VARCHAR2	☒	Categorical	☐
☒ MAT141_SON_H...	VARCHAR2	☒	Categorical	☐
☒ MAT141_TEKRAR...	NUMBER	☒	Numerical	☐
☒ MAT142_ILK_HA...	VARCHAR2	☒	Categorical	☐
☒ MAT142_SON_H...	VARCHAR2	☒	Categorical	☐
☒ MAT142_TEKRAR...	NUMBER	☒	Numerical	☐
☒ MAT175_ILK_HA...	VARCHAR2	☒	Categorical	☐
☒ MAT175_SON_H...	VARCHAR2	☒	Categorical	☐
☒ MAT175_TEKRAR...	NUMBER	☒	Numerical	☐
☒ MEZUNİYET_BAS...	NUMBER	☐	Numerical	☒
☒ OGRENCI_NO	NUMBER	☐	Numerical	☐
☒ OSYM_BASARI_S...	NUMBER	☒	Numerical	☐
☒ OSYM_GIRIS_TURU	NUMBER	☒	Numerical	☐
☒ REGULERLIK	NUMBER	☒	Numerical	☐
☒ TEKRAR_ARTIM_...	FLOAT	☒	Numerical	☐
☒ TEKRAR_ARTIM_...	NUMBER	☒	Numerical	☐
☒ TEKRAR_AZALIM...	FLOAT	☒	Numerical	☐
☒ TEKRAR_AZALIM...	NUMBER	☒	Numerical	☐

Şekil 6.12: Gelişmiş ayarlar

Ayrıca “Classification” düğümü için “Property Inspector”da “Test” ayarlarında model oluşturmada kullanılan veri üzerinde test yapılması istenmediğinden “Perform Test” seçeneğinden işaret kaldırılmıştır (Şekil 6.13).

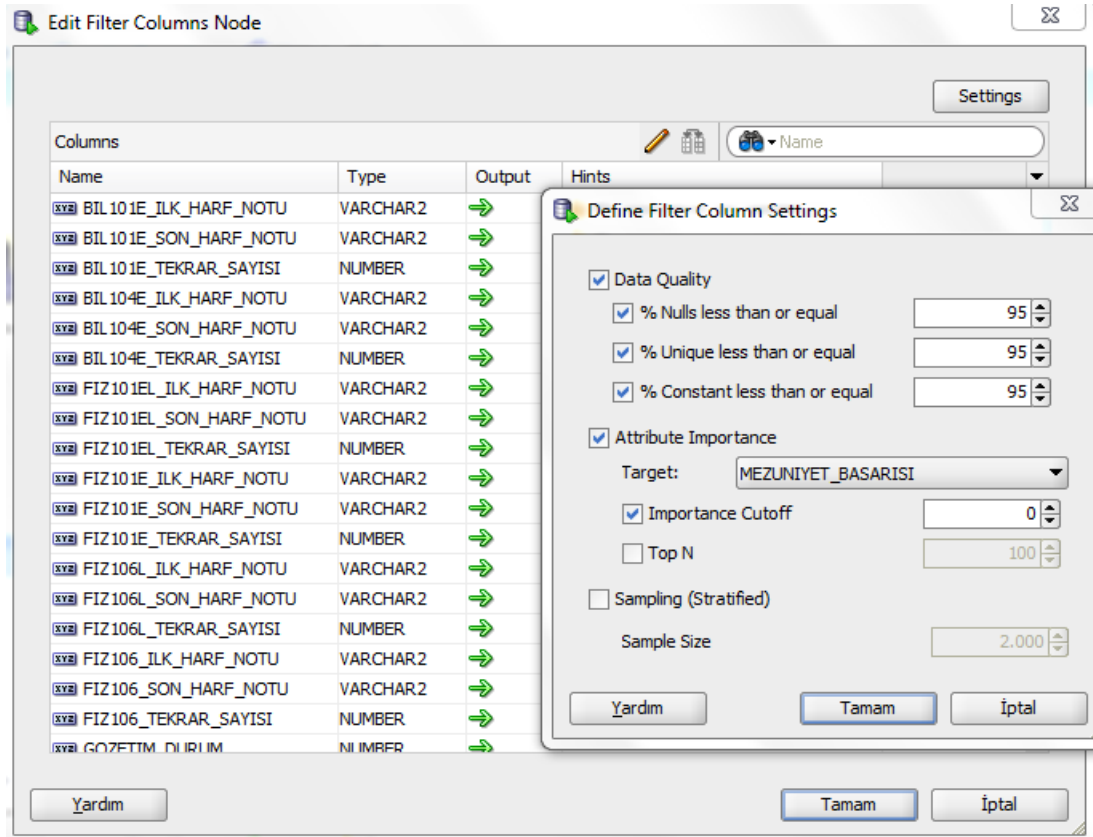


Şekil 6.13: Model oluşturmada test ayarları

Yapılan bu ayarlar ile model oluşumu tamamlanmıştır ve model çalıştırılmıştır.

6.2.1.4 Modelde etkin tahmin edicilerin belirlenmesi (AI)

Bu bölümde tablo içinde yer alan verilerin tahminin başarısına olan etkisini görmek için “Attribute Importance” işleminin uygulanma aşamaları anlatılmıştır [10]. İş akış diyagramına “Component Palette” kısmında “Transforms” altında yer alan “Filter Columns” düğümü eklenmiş ve modelin oluşturulmasında kullanılan “BIRINCI_SINIF_NB” tablosu yüklenmiş veri kaynağı ile bağlantısı sağlanmıştır. “Filter Columns” düğümü ayarları Şekil 6.14’te görüldüğü gibi düzenlenmiştir. “Data Quality” ve “Attribute Importance” seçilmiş, “Top N” seçeneği kaldırılmıştır. “Target” yani hedef “MEZUNİYET_BASARISI” seçilmiştir.



Şekil 6.14: “Attribute Importance” ayarları

Seçilen ayarlarla düğüm çalıştırılmıştır. “Filter Columns” düğümü üzerinde “View Data” seçilerek tahmin edici bilgilerin, tahmin başarısını nasıl ve ne kadar çok etkileyecekleri Şekil 6.15’deki ekran çıktısında görülmektedir.

Attribute Importance			
Data Columns SQL			
Target: MEZUNİYET_BASARISI			
Attribute Ranking			
Name	Type	Rank	Importance
MAT142_ILK_HARF_NOTU	VARCHAR2	1	0,2657
MAT142_SON_HARF_NOTU	VARCHAR2	1	0,2657
MAT112_ILK_HARF_NOTU	VARCHAR2	2	0,1573
MAT112_SON_HARF_NOTU	VARCHAR2	2	0,1573
FIZ106_ILK_HARF_NOTU	VARCHAR2	3	0,121
FIZ106_SON_HARF_NOTU	VARCHAR2	3	0,121
MAT141_ILK_HARF_NOTU	VARCHAR2	4	0,1208
MAT141_SON_HARF_NOTU	VARCHAR2	4	0,1208
MAT142_TEKRAR_SAYISI	NUMBER	5	0,1039
REGULAR_OLMA	NUMBER	6	0,1039
FIZ106L_ILK_HARF_NOTU	VARCHAR2	7	0,0921
FIZ106L_SON_HARF_NOTU	VARCHAR2	7	0,0921
FIZ106L_TEKRAR_SAYISI	NUMBER	7	0,0921
KIM101E_SON_HARF_NOTU	VARCHAR2	8	0,0901
KIM101E_ILK_HARF_NOTU	VARCHAR2	9	0,073
MAT111_ILK_HARF_NOTU	VARCHAR2	10	0,0717
MAT111_SON_HARF_NOTU	VARCHAR2	11	0,0708
FIZ106_TEKRAR_SAYISI	NUMBER	12	0,0602
GOZETIM_PUANI	NUMBER	13	0,052
IYI_DURUM	NUMBER	13	0,052
GOZETIM_DURUM	NUMBER	14	0,0487
MAT175_ILK_HARF_NOTU	VARCHAR2	15	0,0475
FIZ101E_ILK_HARF_NOTU	VARCHAR2	16	0,0445
MAT175_SON_HARF_NOTU	VARCHAR2	17	0,0351
MAT112_TEKRAR_SAYISI	NUMBER	18	0,0336

Şekil 6.15: “BIRINCI_SINIF_NB” AI sonucu

“BIRINCI_SINIF_NB” AI sonucu elde edilen özelliklerin önemlerinin yüzde karşılıkları Tablo 6.2’de verilmiştir. Sonuçlar Bölüm 6.2.1.7’de değerlendirilecektir.

Tablo 6.2: “BIRINCI_SINIF_NB” tablosu için özelliklerin önem yüzdeleri

SÜTUN ADI	DERECE	IMPORTANCE(%)
MAT142_ILK_HARF_NOTU	1	10,54879
MAT142_SON_HARF_NOTU	1	10,54879
MAT112_ILK_HARF_NOTU	2	6,245928
MAT112_SON_HARF_NOTU	2	6,245928
FIZ106_ILK_HARF_NOTU	3	4,804236
FIZ106_SON_HARF_NOTU	3	4,804236
MAT141_ILK_HARF_NOTU	4	4,796987
MAT141_SON_HARF_NOTU	4	4,796987
MAT142_TEKRAR_SAYISI	5	4,125647
REGULAR_OLMA	6	4,123782

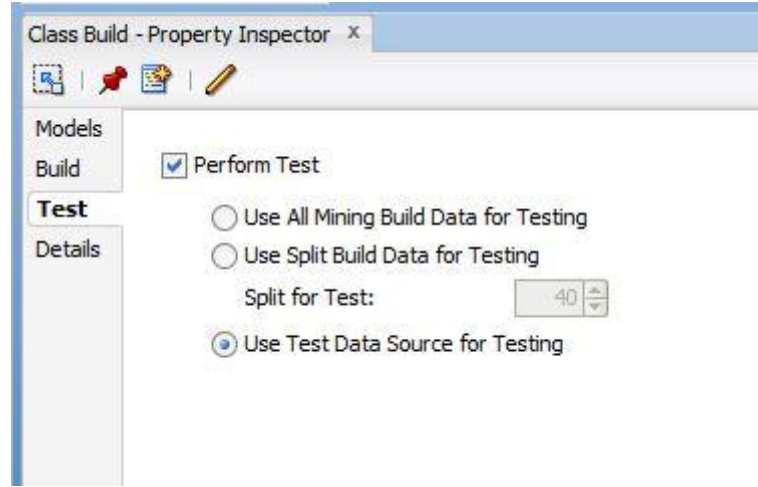
FIZ106L_ILK_HARF_NOTU	7	3,656165
FIZ106L_SON_HARF_NOTU	7	3,656165
FIZ106L_TEKRAR_SAYISI	7	3,656165
KIM101E_SON_HARF_NOTU	8	3,576744
KIM101E_ILK_HARF_NOTU	9	2,900021
MAT111_ILK_HARF_NOTU	10	2,84601
MAT111_SON_HARF_NOTU	11	2,809551
FIZ106_TEKRAR_SAYISI	12	2,391802
GOZETIM_PUANI	13	2,064771
IYI_DURUM	13	2,064771
GOZETIM_DURUM	14	1,932551
MAT175_ILK_HARF_NOTU	15	1,887974
FIZ101E_ILK_HARF_NOTU	16	1,768458
MAT175_SON_HARF_NOTU	17	1,392847
MAT112_TEKRAR_SAYISI	18	1,33282
MAT175_TEKRAR_SAYISI	19	0,561254
ING102_TEKRAR_SAYISI	20	0,460619

“BIRINCI_SINIF_NB” verilerinden sadece en etkili 3 kolon “MAT142_ILK_HARF_NOTU”, “MAT142_SON_HARF_NOTU”, “MAT112_ILK_HARF_NOTU” tahmin edici olarak belirlenerek model tekrar oluşturulmuştur. Modelin test sonucuna göre tahminin doğruluğu %56.36’da kalmıştır. Kurulan modelde bu sütunların etkisi yüksek çıksa da yüksek olasılıkta bir tahmin yapmak için modelin sadece bu sütunları girdi olarak alması yeterli değildir.

6.2.1.5 Modelin testi

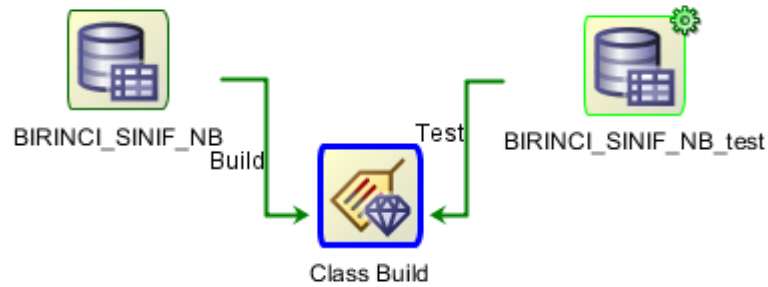
“BIRINCI_SINIF_NB” tablosundaki veriler ile oluşturulan modelin, veri azlığı nedeniyle aynı veriler üzerinde test edilmesi amaçlandığından iş akış diyagramına eklenen “data source” düğümüne yine “BIRINCI_SINIF_NB” tablosu yüklenmiştir, veri kaynağının ismi “BIRINCI_SINIF_NB_test” olarak değiştirilmiştir ve çalıştırılan “Classification” düğümüne bağlanmıştır. Veri kaynağının “Classification” düğümüne bağlandığında “Property Inspector”ın “Test” sekmesinde otomatik olarak

“Perform Test” ayarı altında “Use Test Data Source for Testing”in seçili geldiği görülmektedir (Şekil 6.16).



Şekil 6.16: Test için “Property Inspector” ayarı

“Classification” düğümü çalıştırıldığında Şekil 6.17’de görüldüğü gibi model oluşumu için kullanılan verinin “Classification” düğümü ile bağlantısında “Build”, test için kullanılan verinin “Classification” düğümü ile bağlantısında “Test” ibaresi çıktığı görülmektedir.



Şekil 6.17: “Test” aşaması

“Classification” düğümünün çalışması tamamlandığında “View Test Results” bölümünden testin sonuçları görülür. “Performans Matrix” sekmesinde “Overall Accuracy” modelin doğruluğunu göstermektedir. Şekil 6.18’den de görüldüğü gibi modelin doğruluğu %93.818’dir. Ayrıca sadece 17 öğrenci için mezuniyet başarıları yanlış tahmin edilmiştir.

Performance

Performance Matrix

Lift

Profit

Average Accuracy: 94,6753

Overall Accuracy: 93,8182

Target Value	Total Case Count	Correct Predictions %	Cost	Cost %
0	97	90,7216		
1	61	98,3607		
2	75	92		
3	42	97,619		

▲▼

Performance Matrix: Rows=Actual, Columns=Predicted: ☒ Show totals and cost:

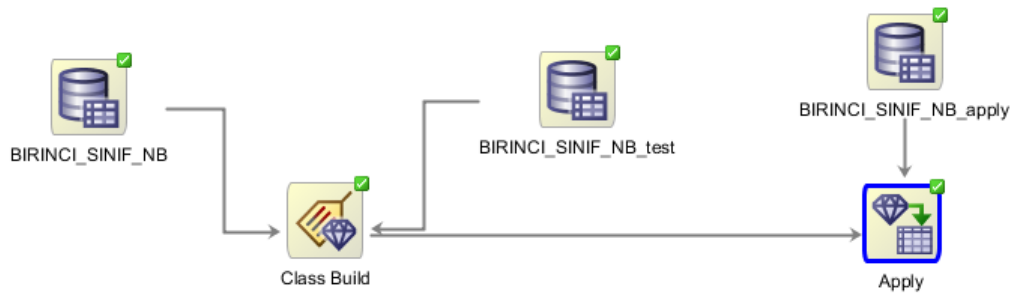
	0	1	2	3	Total	Correct %	Cost
0	88	3	4	2	97	90,7216	
1	1	60	0	0	61	98,3607	
2	4	0	69	2	75	92	
3	1	0	0	41	42	97,619	
Total	94	63	73	45	275		
Correct %	93,617	95,2381	94,5205	91,1111			
Cost							

Şekil 6.18: “BIRINCI_SINIF_NB” için test sonucu

6.2.1.6 Modelin uygulanması

Bu aşamada model oluşturma ve test etmede kullanılan aynı veri tablosu yeni veri kaynağına yüklenerek “BIRINCI_SINIF_NB_apply” şeklinde isimlendirilmiştir. “BIRINCI_SINIF_NB_apply” verisine “Apply” işlemi uygulanmış ve veri içinde yer alan özelliklere göre her öğrenci için öğrencinin mezuniyet başarısı tahmini yapılmıştır.

“Apply” işleminin uygulanması için “Component Palette” de “Evaluate and Apply” sekmesi altında yer alan “Apply” düğümü iş akış diyagramına alındıktan sonra, hem “Class Build” düğümü hem de “BIRINCI_SINIF_NB_apply” veri kaynağı ile Şekil 6.19’deki gibi bağlanmıştır.



Şekil 6.19: “Apply” aşaması

Daha sonra “Apply” düğümünün ayarları düzenlenmiştir. “Data Cols” sekmesinden tüm sütunlar “Selected Attributes” kısmına geçirilip düğüm çalıştırılmıştır. “Apply” düğümünde “View Data” seçilerek verideki gerçek değerleri “MEZUNİYET_BASARISI” sütunu altında, tahmin değerleri ise “CLAS_NB_1_45_PRED” adlı sütun altında, tahmin olasılığı ise “CLAS_NB_1_45_PROB” adlı sütun altında Şekil 6.20’deki gibi gözlemlenmiştir.

OGRENCI_NO	MEZUNİYET_BASARISI	CLAS_NB_1_45_PRED	CLAS_NB_1_45_PROB	TEKRAR_ARTIM_SAYI	KIM101E_ILK_HARF_NOTU	BIL101E_TKRAR_SAYISI	MAT112_TKRAR_SAYISI	FIZ106L_ILK_HARF_NOTU	KIM10
1	8	3	3	1	0	AA	1	1	AA
2	9	0	0	1	0	AA	1	0	YYY
3	10	3	3	1	0	BB	1	1	AA
4	11	0	0	1	0	CB	1	1	BA
5	12	0	3	0,6327	0	BA	1	1	AA
6	13	2	2	0,9969	0	AA	1	1	AA
7	14	1	1	1	0	CB	1	1	AA
8	15	3	3	1	0	BA	1	1	AA
9	16	0	0	1	3	FF	1	1	YYY
10	17	1	1	1	0	CB	1	1	AA
11	18	1	1	1	0	DC	1	1	AA
12	19	0	0	1	0	CB	1	1	BA
13	20	0	0	1	0	CB	1	0	YYY
14	21	2	2	1	0	BB	1	1	AA
15	22	2	2	1	0	BB	1	1	AA
16	23	0	0	1	0	DC	1	0	YYY
17	24	0	0	0,9973	0	CC	1	1	BA
18	25	0	1	0,9239	3	DC	1	1	AA
19	26	0	0	0,9846	0	DC	1	1	BA
20	27	1	1	1	0	AA	1	1	AA
21	28	0	0	1	0	CC	1	0	YYY
22	29	0	0	1	0	DD	1	0	YYY
23	30	1	1	1	0	BB	1	1	AA
24	31	3	3	0,9997	0	BA	1	1	BA

Şekil 6.20: “BIRINCI_SINIF_NB” uygulama sonucu

6.2.1.7 Sonuçlar ve tartışma

Şekil 6.18’den de görüldüğü gibi birinci sınıf verileriyle **mezuniyet başarıları tahmin modelinin doğruluğu %93.82 çıkmıştır**. Mezuniyet başarıları gerçekte “0” yani “mezun değil” iken doğru tahmin edilen 88 öğrenci, yanlış olarak “1” tahmin edilen 3 öğrenci, “2” tahmin edilen 4 öğrenci ve “3” tahmin edilen 2 öğrenci vardır. Mezuniyet başarıları gerçekte “1” yani “düşük başarılı mezun” iken doğru tahmin edilen 60 öğrenci, “0” tahmin edilen 1 öğrenci vardır. Mezuniyet başarıları gerçekte “2” yani “orta başarılı mezun” iken doğru tahmin edilen 69 öğrenci, “0” tahmin edilen 4 öğrenci, “3” tahmin edilen 2 öğrenci vardır. Mezuniyet durumu “yüksek başarılı mezun” iken doğru tahmin edilen 41 öğrenci, “0” tahmin edilen 1 öğrenci vardır. 258 öğrencinin mezuniyet başarı durumları doğru, 17 öğrencinin ise mezuniyet başarı durumları yanlış tahmin edilmiştir (Şekil 6.18).

Uygulama sonucu, birinci sınıf verileri ile öğrencilerin mezuniyet başarılarının yüksek olasılık ile tahmin edilebileceğini göstermiştir. Modelden elde edilen veriler ile dönem sonlarında potansiyel zamanında mezuniyet durumuna giremeyen ve potansiyel düşük başarılı mezun öğrenciler belirlenerek bu öğrencilerin danışmanları bilgilendirilebilir.

Mezuniyet başarıları yanlış tahmin edilen öğrencilere bakıldığında, ID'si 122 olan öğrenci için %100 olasılıkla dört yılda mezun olamamıştır tahmini yapıldığı görülür, gerçekte ise öğrenci yüksek başarı ile 4 yılda mezun olmuştur. Ayrıca ID'si 200 olan öğrenci için %97,58 olasılıkla yüksek başarı ile mezun olmuştur tahmini yapılmış olsa da gerçekte öğrencinin yaz okulları dahil edilmeden 10 dönemde mezun olduğu görülmüştür.

Uygulama sonucunda görülen bu tahminlerden anlaşılır ki; birinci sınıftaki durumu ile bölümü yüksek başarı ile bitirebilme ihtimali olmasına rağmen zamanında mezun olamayan veya birinci sınıftaki durumu ile zamanında mezun olamayacak iken yüksek başarı ile mezun olan öğrenciler az da olsa bulunmaktadır.

Tablo 6.2 sonucuna göre tahmin olasılığına %10,54879 ile en etkili özellikler “MAT142_ILK_HARF_NOTU”, “MAT142_SON_HARF_NOTU” özellikleridir. İleriki sınıfların sonuçlarıyla kıyaslandığında birinci sınıf tablosunun ilk 3 kolonunun etkisinin az olduğu, diğer kolonların da tahminlerin başarısında etkili olduğu söylenebilir.

Ayrıca “Regular” olma durumunun öğrencinin mezuniyet başarısını tahmin etmede göreceli olarak etkilidir.

Özellik önem sırasını gösteren Tablo 6.2’de “OSYM_BASARI_SIRASI”, “OSYM_GIRIS_TURU” bilgilerinin yer almaması, “OSYM_BASARI_SIRASI”, “OSYM_GIRIS_TURU” bilgilerinin öğrencinin bölüme girişini sağlayan bilgiler olmasına rağmen, mezuniyet başarısına etkisinin olmadığını göstermektedir. Ayrıca Tablo 6.2’de tekrar edilen derslerle ilgili hiçbir özellik de yer almaz. Birinci sınıfta tekrar edilen derslerin mezuniyet başarısına etkisinin olmadığı söylenebilir. Aynı şekilde Tablo 6.2’de yer almayan BIL101E, BIL104E, FIZ101EL, KIM101EL ve ING101 derslerinin tüm bilgilerinin, KIM101E’nin tekrar sayısı, FIZ101E’nin ilk harf notu dışındaki bilgilerinin, MAT111 VE MAT141 derslerinin tekrar sayısı bilgilerinin tahminlerde etkili olmadığı görülür.

6.2.2 İkinci sınıf verilerine göre mezuniyet başarısının tahmini

Bu bölümde öğrencilerin ikinci sınıf bilgileriyle mezuniyet başarılarının tahmin edilmesi için kurulan modelin oluşturma, test ve uygulama aşamaları anlatılacaktır. Modelde kullanılan veri tablosunun sütun açıklamaları Tablo 6.3'te verilmiştir.

Tablo 6.3: İkinci sınıf tablosunun sütun açıklamaları

SÜTUN NO	SÜTUN ADI	SÜTUN AÇIKLAMASI
1	OGRENCI_NO	Öğrencinin ID'sini belirtir. "8" ile "497" arasında değer alır.
2	MAT213_TEKRAR_SAYISI	MAT213 dersini tekrar etme sayısını belirtir.
3	MAT213_ILK_HARF_NOTU	MAT213 dersinden ilk aldığı harf notunu belirtir. "AA", "BA", "BB", "CB", "CC", "DC", "DD", "FF", "VF", "YYY" değerlerini alır.
4	MAT213_SON_HARF_NOTU	MAT213 dersinden son aldığı harf notunu belirtir. "AA", "BA", "BB", "CB", "CC", "DC", "DD", "FF", "VF", "YYY" değerlerini alır.
5-37	Bu sütunlar aşağıdaki dersler için sırasıyla Sütun 2, Sütun 3 ve Sütun 4'teki gibi adlandırılmıştır: MAT214, MAT232, MAT322E, MAT221, MAT242, MUH211, MUH212, DNK201E, ING201, TUR101, TUR102	Sütunlar için sırasıyla Sütun 2, Sütun 3, Sütun 4'teki Sütun Açıklamaları geçerlidir.
38	BIRINCI_SNF_ORT	Birinci sınıf sonundaki ağırlıklı genel not ortalamasını belirtir.
39	ILK_ARTIM_SAYI	İkinci sınıf boyunca sınıf dersleri dışında ilk kez aldığı ve ortalamasını artırıcı etkiye sahip derslerin sayısını belirtir. Tüm ilk alınan dersler için derslerin kredileri toplamıdır.
40	ILK_ARTIM_NOT	İkinci sınıf boyunca sınıf dersleri dışında ilk kez aldığı ve ortalamasını artırıcı etkiye sahip derslerin puan etkisini belirtir. Tüm ilk alınan dersler için derslerin kredilerinin notun artış miktarı ile çarpımlarının toplamıdır.
41	ILK_AZALIM_SAYI	İkinci sınıf boyunca sınıf dersleri dışında ilk kez aldığı ve ortalamasını azaltıcı etkiye sahip derslerin sayısını belirtir. Tüm ilk alınan dersler için derslerin kredileri toplamıdır.
42	ILK_AZALIM_NOT	İkinci sınıf boyunca sınıf dersleri dışında ilk kez aldığı ve ortalamasını azaltıcı etkiye sahip derslerin puan etkisini belirtir. Tüm ilk alınan dersler için derslerin kredilerinin notun azalış miktarı ile çarpımlarının toplamıdır.
43	TEKRAR_ARTIM_SAYI	İkinci sınıf boyunca tekrar ettiği derslerden notunu yükselttiklerinin sayısını gösterir. Tüm dersler için dersin tekrar sayısı ile dersin kredisinin çarpılmasından elde edilen değerler toplamıdır.

44	TEKRAR_ARTIM_NOT	İkinci sınıf boyunca tekrar ettiği derslerden notunu yükselttiklerinin getirdiği puanı gösterir. Tüm dersler için dersin yükselme notunun dersin kredisi ile çarpılması ile elde edilen değer toplamıdır.
45	TEKRAR_AZALIM_SAYI	İkinci sınıf boyunca tekrar ettiği derslerden notunu düşürdüklerinin sayısını gösterir. Tüm dersler için dersin tekrar sayısı ile dersin kredisinin çarpılmasından elde edilen değerler toplamıdır.
46	TEKRAR_AZALIM_NOT	İkinci sınıf boyunca tekrar ettiği derslerden notunu düşürdüklerinin getirdiği puanı gösterir. Tüm dersler için dersin yükselme notunun dersin kredisi ile çarpılması ile elde edilen değer toplamıdır.
47	IYI_DURUM	Birinci ve ikinci sınıf boyunca akademik durumunun “iyi durum” olduğu dönem sayısını belirtir.
48	GOZETIM_DURUM	Birinci ve ikinci sınıf boyunca akademik durumunun “gözetim durumu” olduğu dönem sayısını belirtir.
49	GOZETIM_PUANI	Akademik durumu belirten puandır. IYI_DURUM sütun bilgisinden GOZETIM_DURUM sütun bilgisinin çıkarılması ile bulunur. “-6” ile “6” arasında değer alır.
50	REGULAR_OLMA	Regular veya Irregular Olma durumunu belirtir. “Regular”:1, “Irregular”:0 değerlerini alır.
51	OSYM_BASARI_SIRASI	ÖSYM tarafından yapılan sınav sonucuna göre başarı sırasını belirtir.
52	OSYM_GIRIS_TURU	ÖSYM tarafından yapılan sınav sonucuna göre okula kabul ediliş durumunu belirtir.
53	MEZUNİYET_BASARISI	Mezuniyet başarı durumunu belirtir. “0”:mezun değil, “1”:düşük başarılı mezun, “2”:orta başarılı mezun, “3”:yüksek başarılı mezun değerlerini alır.

6.2.2.1 Veri tablosunun hazırlanması ve modelin oluşturulması

Tablo 6.3, Bölüm 6.2.1.1’de anlatılan şekilde Ekler bölümündeki fonksiyonların ikinci sınıf için çalıştırılması ile oluşturulmuştur. Tablo 6.1’den farklı olarak Tablo 6.3’te Sütun 38’de birinci sınıf sonu ortalamasının olduğu görülür. Ayrıca ikinci sınıf dersleri dışında ilk kez alınan derslere ait bilgilere, ortalamayı artıracak etkide ise Sütun 39 ve Sütun 40’da, azaltacak etkiye sahip ise Sütun 41 ve Sütun 42’de yer verilmiştir.

Elde edilen tablo Bölüm 6.2.1.2’de anlatıldığı gibi Oracle Veri Tabanı’na aktarılmış ve “İKINCI_SINIF_NB” şeklinde isimlendirilmiştir. İkinci sınıf verilerine göre mezuniyet başarısının tahmini için oluşturulacak modelde “İKINCI_SINIF_NB” tablosu kullanılacaktır (Şekil 6.21).

İKİNCİ_SINIF_NB											
Data Columns SQL											
View Actual Data Sort... Filter: Enter Where Clause											
OGRENCI_NO	BİRİNCİ_SNF_ORT	DNK20 IE_İLK_HARF_NOTU	DNK20 IE_SON_HARF_NOTU	DNK20 IE_TEKRAR_SAYISI	GOZETİM_DURUM	GOZETİM_PUANI	İLK_ARTIM_NOT	İLK_ARTIM_SAYI	İLK_AZALIM_NOT	İLK_AZAL	
1	8	3,47 AA		AA	1	0	4	146,5	38	12	
2	9	3,67 YYY		YYY	0	0	4	68	17	55	
3	10	3,42 BB		BB	1	0	4	108,5	28	31	
4	11	2,09 CB		CB	1	0	4	62,5	20	20,5	
5	12	2,8 DC		DC	1	0	4	53	16	44	
6	13	2,48 CC		CC	1	0	4	60,5	18	34	
7	14	2,22 CB		CB	1	0	5	58	22	22	
8	15	2,83 CB		CB	1	0	5	137	42	13,5	
9	16	1,55 YYY		YYY	0	4	-3	33	9	0	
10	17	2,38 CC		CC	1	0	5	81,5	24	38	
11	18	1,85 DD		DD	1	0	4	62	20	20	
12	19	2,19 YYY		YYY	0	1	4	33,5	11	14,5	
13	20	3,14 CC		CC	1	0	4	36	9	65,5	
14	21	2,75 DC		DC	1	0	6	36	11	66	
15	22	2,8 CC		CC	1	0	4	83	25	31,5	
16	23	2,32 FF		FF	1	0	4	32	9	8,5	
17	24	2,24 FF		FF	1	0	4	42	14	24	
18	25	1,98 FF		FF	1	0	6	81	28	8	
19	26	1,97 DD		DD	1	0	4	43,5	18	18,5	
20	27	2,35 DC		DC	1	0	4	30	10	41	
21	28	1,69 YYY		YYY	0	2	1	28	9	4,5	
22	29	1,73 YYY		YYY	0	3	-2	6	3	3	
23	30	2,55 DC		DC	1	0	4	62	19	35,5	
24	31	2,04 YYY		YYY	0	0	5	88,5	27	14	

Şekil 6.21: “İKİNCİ_SINIF_NB” tablosundan bir görüntü

“İKİNCİ_SINIF_NB” tablosu üzerinde Naive Bayes yöntemi uygulanacaktır.

Bölüm 6.2.1.2’de anlatıldığı gibi model, “İKİNCİ_SINIF_NB” tablosu kullanılarak, “Target”; “MEZUNİYET_BASARISI”, “CaseID” ; “OGRENCI_NO” seçilerek oluşturulmuştur. Bu sütunlar hariç tüm sütunlar girdi olarak verilmiş ve bu sütunlar için “Auto Prep” kısmının seçili olmamasına dikkat edilmiştir.

6.2.2.2 Modelde etkin tahmin edicilerin belirlenmesi (AI)

Tablo içinde yer alan verilerin olasılıklara olan etkisini görmek için “İKİNCİ_SINIF_NB” tablosuna Bölüm 6.2.1.4’te anlatıldığı gibi AI işlemi uygulanmıştır. Uygulamanın sonucu Şekil 6.22’de verilmiştir.

Attribute Importance			
Data Columns SQL			
Target: MEZUNİYET_BASARISI			
Attribute Ranking			
Name	Type	Rank	Importance
BIRINCI_SNF_OR_T	FLOAT	1	0,3309
MAT322E_SON_HARF_NOTU	VARCHAR2	2	0,2862
MAT222_SON_HARF_NOTU	VARCHAR2	3	0,2837
MAT232_SON_HARF_NOTU	VARCHAR2	4	0,2727
MAT232_ILK_HARF_NOTU	VARCHAR2	5	0,2716
MAT222_ILK_HARF_NOTU	VARCHAR2	6	0,2573
MAT322E_ILK_HARF_NOTU	VARCHAR2	7	0,2478
MAT214_SON_HARF_NOTU	VARCHAR2	8	0,2181
MAT214_ILK_HARF_NOTU	VARCHAR2	9	0,2135
MUH212_ILK_HARF_NOTU	VARCHAR2	10	0,1987
MUH212_SON_HARF_NOTU	VARCHAR2	10	0,1987
ILK_ARTIM_NOT	FLOAT	11	0,171
MAT213_ILK_HARF_NOTU	VARCHAR2	12	0,1489
MAT213_SON_HARF_NOTU	VARCHAR2	13	0,1484
MAT222_TKRAR_SAYISI	NUMBER	14	0,1397
MAT221_ILK_HARF_NOTU	VARCHAR2	15	0,1379
MUH212_TKRAR_SAYISI	NUMBER	16	0,1296
IYI_DURUM	NUMBER	17	0,1259
MAT232_TKRAR_SAYISI	NUMBER	18	0,1178
GOZETIM_PUANI	NUMBER	19	0,1163
REGULAR_OLMA	NUMBER	20	0,1039
MUH211_ILK_HARF_NOTU	VARCHAR2	21	0,1005
MAT322E_TKRAR_SAYISI	NUMBER	22	0,0965
MUH211_SON_HARF_NOTU	VARCHAR2	23	0,0951
ILK_ARTIM_SAYI	NUMBER	24	0,0908

Şekil 6.22: IKINCI_SINIF_NB tablosu için AI sonucu

“IKINCI_SINIF_NB” AI sonucu elde edilen özelliklerin önemlerinin yüzde karşılıkları Tablo 6.4’te verilmiştir. Sonuçlar Bölüm 6.2.1.5’te değerlendirilecektir..

Tablo 6.4: “IKINCI_SINIF_NB” tablosu için özelliklerin önem yüzdeleri

SÜTUN ADI	DERECE	IMPORTANCE(%)
BIRINCI_SNF_OR_T	1	6,268096
MAT322E_SON_HARF_NOTU	2	5,421197
MAT242_SON_HARF_NOTU	3	5,375593
MAT232_SON_HARF_NOTU	4	5,166571
MAT232_ILK_HARF_NOTU	5	5,145529
MAT242_ILK_HARF_NOTU	6	4,874282
MAT322E_ILK_HARF_NOTU	7	4,694151
MAT214_SON_HARF_NOTU	8	4,132009
MAT214_ILK_HARF_NOTU	9	4,043821
MUH212_ILK_HARF_NOTU	10	3,763944

MUH212_SON_HARF_NOTU	10	3,763944
ILK_ARTIM_NOT	11	3,239876
MAT213_ILK_HARF_NOTU	12	2,821512
MAT213_SON_HARF_NOTU	13	2,81123
MAT242_TEKRAR_SAYISI	14	2,647247
MAT221_ILK_HARF_NOTU	15	2,612802
MUH212_TEKRAR_SAYISI	16	2,456138
IYI_DURUM	17	2,384505
MAT232_TEKRAR_SAYISI	18	2,231549
GOZETIM_PUANI	19	2,203043
REGULAR_OLMA	20	1,967455
MUH211_ILK_HARF_NOTU	21	1,903586
MAT322E_TEKRAR_SAYISI	22	1,828525
MUH211_SON_HARF_NOTU	23	1,801323
ILK_ARTIM_SAYI	24	1,719508
ILK_AZALIM_NOT	25	1,695972
MAT214_TEKRAR_SAYISI	26	1,63536
GOZETIM_DURUM	27	1,589062
TEKRAR_ARTIM_NOT	28	1,366303
TEKRAR_ARTIM_SAYI	28	1,366303
MAT221_SON_HARF_NOTU	29	1,309159
DNK201E_ILK_HARF_NOTU	30	0,995731
TEKRAR_AZALIM_SAYI	31	0,988904
DNK201E_SON_HARF_NOTU	32	0,799106
MAT221_TEKRAR_SAYISI	33	0,706461
TUR101_ILK_HARF_NOTU	34	0,657506
MAT213_TEKRAR_SAYISI	35	0,497116
ING201_ILK_HARF_NOTU	36	0,479053
ING201_SON_HARF_NOTU	36	0,479053
ING201_TEKRAR_SAYISI	37	0,157473

“İKİNCİ_SINIF_NB” verilerinden sadece en etkili 3 kolon “BİRİNCİ_SNF_ORT”, “MAT322E_SON_HARF_NOTU”, “MAT242_SON_HARF_NOTU” tahmin edici olarak belirlenerek mezuniyet başarısını tahmin eden model tekrar oluşturulmuştur (Şekil 6.23).

Data Usage		Algorithm Settings		Performance Settings	
Data Usage					
Input Ignore					
Attributes	Data Type	Input	Mining Type	Auto Prep	
BIRINCI_SNF_ORT	FLOAT		Numerical		
MAT222_SON_H...	VARCHAR2		Categorical		
MAT322E_SON_...	VARCHAR2		Categorical		
MEZUNİYET_BAS...	NUMBER		Numerical		
OGRENCI_NO	NUMBER		Numerical		

Şekil 6.23: “İKİNCİ_SINIF_NB”nin en etkili 3 özelliğini tahmin edici belirleme

Sadece ilk üç özelliğin alınması ile elde edilen bu modelin test sonucuna göre tahminin doğruluğu %89.4545 çıkmıştır (Şekil 6.24).

Sonuçlar Bölüm 6.2.2.5’te değerlendirilecektir.

Performance

Performance Matrix

Lift

Profit

Average Accuracy: 89,1606

Overall Accuracy: 89,4545

Target Value	Total Case Count	Correct Predictions %	Cost	Cost %
0	97	94,8454		
1	61	83,6066		
2	75	85,3333		
3	42	92,8571		

▲▼

Performance Matrix: Rows=Actual, Columns=Predicted: ☐ Show totals and cost:

	0	1	2	3	
0	92	4	0	1	
1	7	51	3	0	
2	2	6	64	3	
3	0	2	1	39	
Total	101	63	68	43	
Correct %	91,0891	80,9524	94,1176	90,6977	
Cost					

Şekil 6.24: “İKİNCİ_SINIF_NB” için en etkili 3 özellik ile mezuniyet başarısı tahmin etme test sonucu

6.2.2.3 Modelin testi

İkinci sınıf verilerine göre oluşturulan modelin test aşaması Bölüm 6.2.1.3'te anlatılan şekilde gerçekleştirilmiştir. "IKINCI_SINIF_NB_test" veri kaynağında test işlemi yapılmıştır.

Test işlemi tamamlandığında modelin doğruluğunun %100 olduğu görülmektedir. Sekil 6.25'ten de görüldüğü gibi hiç hatalı tahmin yapılmamıştır.

Performance

Performance Matrix

Lift

Profit

Average Accuracy: 100

Overall Accuracy: 100

Target Value	Total Case Count	Correct Predictions %	Cost	Cost %
0	97	100		
1	61	100		
2	75	100		
3	42	100		

▲▼

Performance Matrix: Rows=Actual, Columns=Predicted: ☐ Show totals and cost:

	0	1	2	3	
0	97	0	0	0	
1	0	61	0	0	
2	0	0	75	0	
3	0	0	0	42	
Total	97	61	75	42	
Correct %	100	100	100	100	
Cost					

Şekil 6.25: "IKINCI_SINIF_NB" için test sonucu

6.2.2.4 Modelin uygulanması

Bu aşamada model oluşturma ve test etmede kullanılan aynı veri tablosu yeni veri kaynağına yüklenerek "IKINCI_SINIF_NB_apply" şeklinde isimlendirilmiştir. "IKINCI_SINIF_NB_apply" verisine "Apply" işlemi Bölüm 6.2.1.4'te açıklandığı gibi uygulanmış ve veri içinde yer alan özelliklere göre her öğrenci için öğrencinin mezuniyet başarısı tahmini yapılmıştır.

"Apply" işlemi sonucunda gerçek değerleri "MEZUNİYET_BASARISI" sütunu altında, tahmin değerleri ise "CLAS_NB_1_46_PRED" adlı sütun altında, tahmin olasılığı ise "CLAS_NB_1_46_PROB" adlı sütun altında Şekil 6.26'daki gibi gözlemlenmiştir.

OGRENCI_NO	MEZUNİYET_BASARISI	CLAS_NB_1_46_PRED	CLAS_NB_1_46_PROB	MAT222_TKRAR_SAYISI	MAT232_TKRAR_SAYISI	TEKRAR_ARTIM_SAYI	MAT214_ILK_HARF_NOTU	ING201_TKRAR_SAYISI	MAT222
1	8	3	3	1	1	1	0	BB	1
2	9	0	0	1	0	0	0	YYY	1
3	10	3	3	1	1	1	0	CB	1
4	11	0	0	1	1	1	3	DC	1
5	12	0	0	1	1	1	0	CB	1
6	13	2	2	1	1	1	5	DC	1
7	14	1	1	1	1	1	6	CB	1
8	15	3	3	1	2	1	4	DC	1
9	16	0	0	1	0	0	11	YYY	1
10	17	1	1	1	2	1	4	CB	1
11	18	1	1	1	1	1	3	DD	1
12	19	0	0	1	1	1	8	FF	0
13	20	0	0	1	0	0	0	YYY	1
14	21	2	2	1	1	1	3	CB	1
15	22	2	2	1	1	1	0	BB	1
16	23	0	0	1	1	0	5	YYY	1
17	24	0	0	0,953	1	1	3	FF	1
18	25	0	0	1	1	1	6	DD	0
19	26	0	0	1	1	1	6	DD	1
20	27	1	1	1	1	1	0	DC	1
21	28	0	0	1	0	0	13	YYY	1
22	29	0	0	1	0	0	0	YYY	0
23	30	1	1	1	1	1	0	CC	1
24	31	3	3	1	1	1	11	CB	1

Şekil 6.26: “İKİNCİ_SINIF_NB” uygulama sonucu

6.2.2.5 Sonuçlar ve tartışma

Şekil 6.25’ten de görüldüğü gibi ikinci sınıf verileri ile mezuniyet başarısı tahmin etme **modelinin doğruluğu %100 çıkmıştır**. Şekil 6.25’teki “Performans Matrix” sonucundan modelde hiç yanlış tahmin yapılmadığı görülmektedir. Mezuniyet başarısı gerçekte “0” yani “mezun değil” iken doğru tahmin edilen 97 öğrenci, “1” yani “düşük başarılı mezun” iken doğru tahmin edilen 61 öğrenci, “2” yani “orta başarılı mezun” iken doğru tahmin edilen 75 öğrenci, “3” yani “yüksek başarılı mezun” iken doğru tahmin edilen 42 kişi vardır. Uygulama sonucu, ikinci sınıf verileri ile öğrencilerin mezuniyet başarılarının %100 tahmin edilebileceğini gösterir.

Tablo 6.4 sonucuna göre mezuniyet başarısını tahmin etmede %6,268096 ile en etkili özellik “BİRİNCİ_SNF_ORT” özelliğidir. Öğrencinin öğrenim başarısıyla ilgili net bir veri olduğundan birinci sınıf ortalamasının en etkili özellik çıkması şaşırtıcı değildir.

Bölüm 6.2.2.2’de yapılan en etkin özelliklerle oluşturulan modele göre öğrencinin sadece birinci sınıf ortalaması, MAT322E ve MAT242 dersinden aldığı son harf notu bilgileri öğrencilerin mezuniyet başarılarını yüksek olasılık ile belirlemede yeterlidir. Aynı zamanda MAT322E ve MAT242 derslerinin Bölüm 4’te çıkarılan istatistiklere

göre ortalaması en düşük ilk üç bölüm dersi arasında yer aldıkları görülmektedir (Tablo 4.35).

Tablo 6.4'ten görüldüğü gibi IYI_DURUM, GOZETIM_DURUM, GOZETIM_PUANI ve REGULAR_OLMA bilgilerinin göreceli olarak etkili oldukları görülür.

Tablo 6.4'te yer almayan DNK201_TEKRAR_SAYISI, MUH211_TEKRAR_SAYISI, ILK_AZALIM_SAYI, TEKRAR_AZALIM_NOT özelliklerinin mezuniyet başarılarını tahmin etmede etkilerinin hiç olmadığı görülmektedir. Ayrıca TUR101 dersinin %0.66 öneme sahip ilk harf notu bilgisi dışındaki bilgilerinin ve TUR102 dersinin tüm bilgilerinin mezuniyet başarılarını tahmin etmede hiç etkili olmadığı görülmektedir.

ÖSYM verilerinin ikinci sınıf bilgileri ile mezuniyet başarısı tahmin etme modelinde de mezuniyet başarısını tahmin etmede etkili olmadığı görülür.

6.2.3 Üçüncü sınıf verilerine göre mezuniyet başarısının tahmini

Bu bölümde öğrencilerin üçüncü sınıf bilgileriyle mezuniyet başarılarının tahmin edilmesi için kurulan modelin oluşturma, test ve uygulama aşamaları anlatılacaktır. Modelde kullanılan veri tablosunun sütun açıklamaları Tablo 6.5'te verilmiştir.

Tablo 6.5: Üçüncü sınıf tablosunun sütun açıklamaları

SÜTUN NO	SÜTUN ADI	SÜTUN AÇIKLAMASI
1	OGRENCI_NO	Öğrencinin ID'sini belirtir. "8" ile "497" arasında değer alır.
2	MAT331_TEKRAR_SAYISI	MAT331 dersini tekrar etme sayısını belirtir.
3	MAT331_ILK_HARF_NOTU	MAT331 dersinden ilk aldığı harf notunu belirtir. "AA", "BA", "BB", "CB", "CC", "DC", "DD", "FF", "VF", "YYY" değerlerini alır.
4	MAT331_SON_HARF_NOTU	MAT331 dersinden son aldığı harf notunu belirtir. "AA", "BA", "BB", "CB", "CC", "DC", "DD", "FF", "VF", "YYY" değerlerini alır.
5-25	Bu sütunlar aşağıdaki dersler için sırasıyla Sütun 2, Sütun 3 ve Sütun 4'teki gibi adlandırılmıştır: MAT341E, MAT355, MAT332, MAT342E, MUH311, MUH312, MUH321	Sütunlar için sırasıyla Sütun 2, Sütun 3, Sütun 4'teki Sütun Açıklamaları geçerlidir.
26	IKINCI_SNF_ORT	İkinci sınıf sonundaki ağırlıklı genel not ortalamasını belirtir.

27	ILK_ARTIM_SAYI	Üçüncü sınıf boyunca sınıf dersleri dışında ilk kez aldığı ve ortalamasını artırıcı etkiye sahip derslerin sayısını belirtir. Tüm ilk alınan dersler için derslerin kredileri toplamıdır.
28	ILK_ARTIM_NOT	Üçüncü sınıf boyunca sınıf dersleri dışında ilk kez aldığı ve ortalamasını artırıcı etkiye sahip derslerin puan etkisini belirtir. Tüm ilk alınan dersler için derslerin kredilerinin notun artış miktarı ile çarpımlarının toplamıdır.
29	ILK_AZALIM_SAYI	Üçüncü sınıf boyunca sınıf dersleri dışında ilk kez aldığı ve ortalamasını azaltıcı etkiye sahip derslerin sayısını belirtir. Tüm ilk alınan dersler için derslerin kredileri toplamıdır.
30	ILK_AZALIM_NOT	Üçüncü sınıf boyunca sınıf dersleri dışında ilk kez aldığı ve ortalamasını azaltıcı etkiye sahip derslerin puan etkisini belirtir. Tüm ilk alınan dersler için derslerin kredilerinin notun azalış miktarı ile çarpımlarının toplamıdır.
31	TEKRAR_ARTIM_SAYI	Üçüncü sınıf boyunca tekrar ettiği derslerden notunu yükselttiklerinin sayısını gösterir. Tüm dersler için dersin tekrar sayısı ile dersin kredisinin çarpılmasından elde edilen değerler toplamıdır.
32	TEKRAR_ARTIM_NOT	Üçüncü sınıf boyunca tekrar ettiği derslerden notunu yükselttiklerinin getirdiği puanı gösterir. Tüm dersler için dersin yükselme notunun dersin kredisi ile çarpılması ile elde edilen değer toplamıdır.
33	TEKRAR_AZALIM_SAYI	Üçüncü sınıf boyunca tekrar ettiği derslerden notunu düşürdüklerinin sayısını gösterir. Tüm dersler için dersin tekrar sayısı ile dersin kredisinin çarpılmasından elde edilen değerler toplamıdır.
34	TEKRAR_AZALIM_NOT	Üçüncü sınıf boyunca tekrar ettiği derslerden notunu düşürdüklerinin getirdiği puanı gösterir. Tüm dersler için dersin yükselme notunun dersin kredisi ile çarpılması ile elde edilen değer toplamıdır.
35	IYI_DURUM	Birinci, ikinci ve üçüncü sınıf boyunca akademik durumunun “iyi durum” olduğu dönem sayısını belirtir.
36	GOZETIM_DURUM	Birinci, ikinci ve üçüncü sınıf boyunca akademik durumunun “gözetim durumu” olduğu dönem sayısını belirtir.
37	GOZETIM_PUANI	Akademik durumu belirten puandır. IYI_DURUM sütun bilgisinden GOZETIM_DURUM sütun bilgisinin çıkarılması ile bulunur. “-9” ile “9” arasında değer alır.
38	REGULAR_OLMA	Regular veya Irregular Olma durumunu belirtir. “Regular”:1, “Irregular”:0 değerlerini alır.
39	OSYM_BASARI_SIRASI	ÖSYM tarafından yapılan sınav sonucuna göre başarı sırasını belirtir.
40	OSYM_GIRIS_TURU	ÖSYM tarafından yapılan sınav sonucuna göre okula kabul ediliş durumunu belirtir.
41	MEZUNİYET_BASARISI	Mezuniyet başarı durumunu belirtir. “0”:mezun değil, “1”:düşük başarılı mezun, “2”:orta başarılı mezun, “3”:yüksek başarılı mezun değerlerini alır.

6.2.3.1 Veri tablosunun hazırlanması ve modelin oluşturulması

Tablo 6.5, Bölüm 6.2.1.1’de anlatılan şekilde Ekler bölümündeki fonksiyonların üçüncü sınıf için çalıştırılması ile oluşturulmuştur. Tablo 6.1’den farklı olarak Tablo 6.5’te Sütun 26’da ikinci sınıf sonu ortalamasının olduğu görülür. Ayrıca üçüncü sınıf dersleri dışında ilk kez alınan derslere ait bilgilere, ortalamayı artıracak etkide ise Sütun 27 ve Sütun 28’de, azaltacak etkiye sahip ise Sütun 29 ve Sütun 30’da yer verilmiştir.

Elde edilen tablo Bölüm 6.2.1.2’de anlatıldığı gibi Oracle Veri Tabanı’na aktarılmış ve “UCUNCU_SINIF_NB” şeklinde isimlendirilmiştir. Üçüncü sınıf verilerine göre mezuniyet başarısının tahmini için oluşturulacak modelde “UCUNCU_SINIF_NB” tablosu kullanılacaktır (Şekil 6.27).

OGRENCI_NO	MUH311_ILK_HARF_NOTU	TEKRAR_ARTIM_SAYI	MUH321_ILK_HARF_NOTU	MAT331_SON_HARF_NOTU	ILK_ARTIM_NOT	MAT342E_TKRAR_SAYISI	MAT342E_SON_HARF_NOTU	MAT341E_SON_HARF_NO
1	8	BA	0 BA		CB	36	1 AA	
2	9	YYY	0 YYY		FF	28	0 YYY	
3	10	AA	3 BB		BB	40	1 CB	
4	11	DD	8 DD		CC	9	1 FF	
5	12	DC	8 DC		CC	40	1 CC	
6	13	CB	16 YYY		CB	38	1 DC	
7	14	CB	3 CC		FF	59	1 BB	
8	15	AA	7 BB		DC	55,5	1 AA	
9	16	FF	5 YYY		YYY	56,5	0 YYY	
10	17	DD	0 BB		DC	53,5	1 CB	
11	18	FF	14 YYY		CB	33	1 FF	
12	19	FF	24 DD		DC	21	0 YYY	
13	20	DC	0 CB		CC	45	0 YYY	
14	21	CC	6 BB		FF	83	1 DC	
15	22	DC	0 DC		FF	56	1 CB	
16	23	YYY	6 FF		YYY	23	0 YYY	
17	24	DD	14 YYY		DC	21	1 DD	
18	25	DC	3 BB		FF	40,5	1 DD	
19	26	FF	7 DC		DC	15	1 DC	
20	27	CB	7 CB		DD	39	1 DD	
21	28	CC	15 BB		DD	84,5	0 YYY	
22	29	YYY	18 YYY		YYY	18,5	0 YYY	
23	30	CB	11 BB		DD	46,5	1 DC	
24	31	CB	2 BB		CC	65,5	1 AA	

Şekil 6.27: “UCUNCU_SINIF_NB” tablosundan bir görüntü

“UCUNCU_SINIF_NB” tablosu üzerinde Naive Bayes yöntemi uygulanacaktır. Bölüm 6.2.1.2’de anlatıldığı gibi model, “UCUNCU_SINIF_NB” tablosu kullanılarak, “Target”; “MEZUNİYET_BASARISI”, “CaseID” ; “OGRENCI_NO” seçilerek oluşturulmuştur. Bu sütunlar hariç tüm sütunlar girdi olarak verilmiş ve bu sütunlar için “Auto Prep” kısmının seçili olmamasına dikkat edilmiştir.

6.2.3.2 Modelde etkin tahmin edicilerin belirlenmesi (AI)

Tablo içinde yer alan verilerin olasılıklara olan etkisini görmek için “UCUNCU_SINIF_NB” tablosuna Bölüm 6.2.1.4’te anlatıldığı gibi AI işlemi uygulanmıştır. Uygulamanın sonucu Şekil 6.28’de verilmiştir.

Attribute Importance			
Data Columns SQL			
Target: MEZUNİYET_BASARISI			
Attribute Ranking			
Name	Type	Rank	Importance
İKİNCİ_SNF_ORT	FLOAT	1	0,4651
MAT342E_ILK_HARF_NOTU	VARCHAR2	2	0,3376
MUH312_ILK_HARF_NOTU	VARCHAR2	3	0,3284
MAT342E_SON_HARF_NOTU	VARCHAR2	4	0,3209
MUH312_SON_HARF_NOTU	VARCHAR2	5	0,2779
MAT332_ILK_HARF_NOTU	VARCHAR2	6	0,2716
MAT332_SON_HARF_NOTU	VARCHAR2	7	0,2527
MUH311_ILK_HARF_NOTU	VARCHAR2	8	0,2424
MUH311_SON_HARF_NOTU	VARCHAR2	9	0,2345
MAT341E_SON_HARF_NOTU	VARCHAR2	10	0,2054
ILK_ARTIM_NOT	FLOAT	11	0,2022
MUH312_TEKRAR_SAYISI	NUMBER	12	0,2015
MAT341E_ILK_HARF_NOTU	VARCHAR2	13	0,2009
MAT342E_TEKRAR_SAYISI	NUMBER	14	0,1819
MAT331_ILK_HARF_NOTU	VARCHAR2	15	0,1777
MAT331_SON_HARF_NOTU	VARCHAR2	16	0,1761
GOZETIM_PUANI	NUMBER	17	0,1494
MAT332_TEKRAR_SAYISI	NUMBER	18	0,1317
ILK_ARTIM_SAYI	NUMBER	19	0,1316
GOZETIM_DURUM	NUMBER	20	0,1309
TYI_DURUM	NUMBER	21	0,1215
MAT355_ILK_HARF_NOTU	VARCHAR2	22	0,1182
MAT355_SON_HARF_NOTU	VARCHAR2	22	0,1182
REGULAR_OLMA	NUMBER	23	0,1039
MUH321_ILK_HARF_NOTU	VARCHAR2	24	0,0847

Şekil 6.28: UCUNCU_SINIF_NB tablosu için AI sonucu

“UCUNCU_SINIF_NB” AI sonucu elde edilen özelliklerin önemlerinin yüzde karşılıkları Tablo 6.6’da verilmiştir. Sonuçlar Bölüm 6.2.1.5’te değerlendirilecektir.

Tablo 6.6: “UCUNCU_SINIF_NB” tablosu için özelliklerin önem yüzdeleri

SÜTUN ADI	DERECE	IMPORTANCE(%)
İKİNCİ_SNF_ORT	1	8,197799
MAT342E_ILK_HARF_NOTU	2	5,950072
MUH312_ILK_HARF_NOTU	3	5,78737
MAT342E_SON_HARF_NOTU	4	5,655525
MUH312_SON_HARF_NOTU	5	4,898086
MAT332_ILK_HARF_NOTU	6	4,787509
MAT332_SON_HARF_NOTU	7	4,454356
MUH311_ILK_HARF_NOTU	8	4,271799
MUH311_SON_HARF_NOTU	9	4,133619
MAT341E_SON_HARF_NOTU	10	3,620304

ILK_ARTIM_NOT	11	3,563829
MUH312_TEKRAR_SAYISI	12	3,550917
MAT341E_ILK_HARF_NOTU	13	3,541409
MAT342E_TEKRAR_SAYISI	14	3,205814
MAT331_ILK_HARF_NOTU	15	3,1324
MAT331_SON_HARF_NOTU	16	3,104237
GOZETIM_PUANI	17	2,632835
MAT332_TEKRAR_SAYISI	18	2,321368
ILK_ARTIM_SAYI	19	2,319091
GOZETIM_DURUM	20	2,307562
IYI_DURUM	21	2,142008
MAT355_ILK_HARF_NOTU	22	2,083619
MAT355_SON_HARF_NOTU	22	2,083619
REGULAR_OLMA	23	1,830404
MUH321_ILK_HARF_NOTU	24	1,492047
ILK_AZALIM_NOT	25	1,455808
MAT331_TEKRAR_SAYISI	26	1,376425
TEKRAR_AZALIM_SAYI	27	1,27517
MUH321_SON_HARF_NOTU	28	1,227494
TEKRAR_ARTIM_SAYI	29	1,117746
MAT341E_TEKRAR_SAYISI	30	0,766254
TEKRAR_ARTIM_NOT	31	0,7473
MUH321_TEKRAR_SAYISI	32	0,403231
MAT355_TEKRAR_SAYISI	33	0,335526
MUH311_TEKRAR_SAYISI	34	0,22745

“UCUNCU_SINIF_NB” verilerinden sadece en etkili 3 kolon “İKİNCİ_SNF_ORT”, “MAT342E_ILK_HARF_NOTU”, “MUH312_SON_HARF_NOTU” tahmin edici olarak belirlenerek mezuniyet başarısını tahmin eden model tekrar oluşturulmuştur (Şekil 6.29).

Attributes	Data Type	Input	Mining Type	Auto Prep
IKINCI_SNF_ORT	FLOAT	☒	Numerical	☐
MAT342E_ILK_H...	VARCHAR2	☒	Categorical	☐
MEZUNİYET_BAS...	NUMBER	☐	Numerical	☒
MUH312_ILK_HA...	VARCHAR2	☒	Categorical	☐
OGRENCI_NO	NUMBER	☐	Numerical	☐

Şekil 6.29: “UCUNCU_SINIF_NB”nin en etkili 3 özelliğini tahmin edici belirleme

Modelin test sonucuna göre tahminin doğruluğu %90.5455 çıkmıştır (Şekil 6.30).

Sonuçlar Bölüm 6.2.2.5’te değerlendirilecektir.

Performance

Performance Matrix

Lift

Profit

Average Accuracy: 90,8365

Overall Accuracy: 90,5455

Target Value	Total Case Count	Correct Predictions %	Cost	Cost %
0	97	93,8144		
1	61	85,2459		
2	75	86,6667		
3	42	97,619		

▲▼

Performance Matrix: Rows=Actual, Columns=Predicted: ☐ Show totals and cost:

	0	1	2	3	
0	91	3	3	0	
1	3	52	6	0	
2	2	7	65	1	
3	0	1	0	41	
Total	96	63	74	42	
Correct %	94,7917	82,5397	87,8378	97,619	
Cost					

Şekil 6.30: “UCUNCU_SINIF_NB” için en etkili 3 özellik ile mezuniyet başarısı tahmin etme test sonucu

6.2.3.3 Modelin testi

Üçüncü sınıf verilerine göre oluşturulan modelin test aşaması Bölüm 6.2.1.3’te anlatılan şekilde gerçekleştirilmiştir. “UCUNCU_SINIF_NB_test” veri kaynağında test işlemi yapılmıştır

Test işlemi tamamlandığında **modelin doğruluğunun %99,2727 olduğu görülmektedir**. Şekil 6.31’den de görüldüğü gibi 2 adet hatalı tahmin yapılmıştır.

Performance

Performance Matrix

Lift

Profit

Average Accuracy: 99,4089

Overall Accuracy: 99,2727

Target Value	Total Case Count	Correct Predictions %	Cost	Cost %
0	97	98,9691		
1	61	100		
2	75	98,6667		
3	42	100		

▲▼

Performance Matrix: Rows=Actual, Columns=Predicted: ☐ Show totals and cost:

	0	1	2	3
0	96	0	0	1
1	0	61	0	0
2	1	0	74	0
3	0	0	0	42
Total	97	61	74	43
Correct %	98,9691	100	100	97,6744
Cost				

Şekil 6.31: “UCUNCU_SINIF_NB” için test sonucu

6.2.3.4 Modelin uygulanması

Bu aşamada model oluşturma ve test etmede kullanılan aynı veri tablosu yeni veri kaynağına yüklenerek “UCUNCU_SINIF_NB_apply” şeklinde isimlendirilmiştir. “UCUNCU_SINIF_NB_apply” verisine “Apply” işlemi Bölüm 6.2.1.4’te açıklandığı gibi uygulanmış ve veri içinde yer alan özelliklere göre her öğrenci için öğrencinin mezuniyet başarısı tahmini yapılmıştır.

“Apply” işlemi sonucunda gerçek değerleri “MEZUNİYET_BASARISI” sütunu altında, tahmin değerleri ise “CLAS_NB_1_47_PRED” adlı sütun altında, tahmin olasılığı ise “CLAS_NB_1_47_PROB” adlı sütun altında Şekil 6.32’deki gibi gözlemlenmiştir.

OGRENCI_NO	MEZUNİYET_BASARISI	CLAS_NB_1_47_PRED	CLAS_NB_1_47_PROB	MUH311_ILK_HARF_NOTU	TEKRAR_ARTIM_SAYI	MUH321_ILK_HARF_NOTU	MAT331_SON_HARF_NOTU	ILK_ARTIM_NOTU	MAT342E_TE
1	8	3	3	1	BA	0	BA	CB	36
2	9	0	0	1	YYY	0	YYY	FF	28
3	10	3	3	1	AA	3	BB	BB	40
4	11	0	0	1	DD	8	DD	CC	9
5	12	0	0	0,999	DC	8	DC	CC	40
6	13	2	2	1	CB	16	YYY	CB	38
7	14	1	1	1	CB	3	CC	FF	59
8	15	3	3	1	AA	7	BB	DC	55,5
9	16	0	0	1	FF	5	YYY	YYY	56,5
10	17	1	1	1	DD	0	BB	DC	53,5
11	18	1	1	1	FF	14	YYY	CB	33
12	19	0	0	1	FF	24	DD	DC	21
13	20	0	0	1	DC	0	CB	CC	45
14	21	2	2	1	CC	6	BB	FF	83
15	22	2	2	1	DC	0	DC	FF	56
16	23	0	0	1	YYY	6	FF	YYY	23
17	24	0	0	1	DD	14	YYY	DC	21
18	25	0	0	0,9998	DC	3	BB	FF	40,5
19	26	0	0	1	FF	7	DC	DC	15
20	27	1	1	1	CB	7	CB	DD	39
21	28	0	0	1	CC	15	BB	DD	84,5
22	29	0	0	1	YYY	18	YYY	YYY	18,5
23	30	1	1	1	CB	11	BB	DD	46,5
24	31	3	3	1	CB	2	BB	CC	65,5

Şekil 6.32: “UCUNCU_SINIF_NB” uygulama sonucu

6.2.3.5 Sonuçlar ve tartışma

Şekil 6.31’den de görüldüğü gibi üçüncü sınıf verileri ile **mezuniyet başarısı tahmin etme modelinin doğruluğu %99,27 çıkmıştır**. “Performans Matrix”ten modelde 273 öğrenci için doğru tahmin yapılırken, 2 öğrenci için yanlış tahmin yapıldığı görülmektedir. Mezuniyet başarısı gerçekte “0” yani “mezun değil” iken doğru tahmin edilen 96 öğrenci, “3” tahmin edilen 1 öğrenci vardır. “1” yani “düşük başarılı mezun” iken doğru tahmin edilen 61 öğrenci vardır. “2” yani “orta başarılı mezun” iken doğru tahmin edilen 74 öğrenci, “0” tahmin edilen 1 öğrenci vardır. “3” yani “yüksek başarılı mezun” iken doğru tahmin edilen 42 öğrenci vardır. Uygulama sonucu, üçüncü sınıf verileri ile öğrencilerin mezuniyet başarılarının yüksek olasılık ile tahmin edilebileceğini gösterir.

Modelin uygulama sonucundan, ID’si 200 olan öğrenci için %100 olasılıkla yüksek başarı ile mezun olmuştur tahmini yapılmış olsa da gerçekte öğrencinin dört yılda mezun olamadığı görülür. 200 ID’li öğrencinin dönemlere göre ortalamalarına bakıldığında öğrencinin öğrenimini yaz okulları hariç 11 yarıyılta tamamladığı ve tüm dönemlerde ortalamasının en az 3.39 olduğu görülmektedir (Şekil 6.33). Bu durumundan dolayı 200 ID’li öğrenci istisna olarak sonuçlarda görülmektedir.

	OGRENCI_NO	DONEM	DONEM_GPA	TOPLAM_GPA
1	200	200810	3.45	3.45
2	200	200820	3.30	3.39
3	200	200910	3.52	3.44
4	200	200920	3.74	3.59
5	200	201010	3.67	3.61
6	200	201020	3.69	3.62
7	200	201030	3.57	3.62
8	200	201110	3.65	3.63
9	200	201120	3.78	3.65
10	200	201210	3.30	3.62
11	200	201220	3.40	3.61

Şekil 6.33: 200 ID’li öğrencinin dönemlere göre ortalaması

Tablo 6.6 sonucuna göre tahmin olasılığına %8,197799 etki ederek en etkili özellik “İKINCI_SNF_ORT” özelliğidir. Tablo 6.4’ün sonucunda birinci sınıf ortalamasının en etkili özellik olarak çıkmasına benzer şekilde ikinci sınıf ortalamasının en etkili özellik çıkması şaşırtıcı değildir.

Bölüm 6.2.3.2’de yapılan en etkin özelliklerle oluşturulan modele göre öğrencinin sadece ikinci sınıf ortalaması, MAT342E dersinden ve MUH312 dersinden ilk aldığı harf notu bilgileri öğrencilerin mezuniyet başarılarını yüksek olasılık ile belirlemede yeterlidir.

ÖSYM verilerinin üçüncü sınıf bilgileri ile mezuniyet başarıları tahmin etme modelinde de mezuniyet başarılarını tahmin etmede etkili olmadığı görülür.

Tablo 6.6’da yer almayan ILK_AZALIM_SAYI, TEKRAR_AZALIM_NOT özelliklerinin mezuniyet başarılarını tahmin etmede etkilerinin hiç olmadığı görülmektedir.

6.3. Birliktelik Kuralları Uygulaması

Bu bölümde İstanbul Teknik Üniversitesi Öğrenci İşleri Daire Başkanlığı’ndan elde edilen Matematik Mühendisliği mezun öğrenci not döküm verileri kullanılarak öğrencilerin bölüm dersleri başarı durumları arasındaki ilişkiler üzerine kurulan modelin oluşturma ve sonuç kısımları anlatılmaktadır.

Öğrencilerin kurulan bu model kapsamında ilişkileri çıkarılacak bölüm dersleri: “MAT111, MAT112, MAT141, MAT142, MAT175, MAT213, MAT214, MAT221,

MAT232, MAT242, MAT322E, MAT331, MAT332, MAT355, MAT341E, MAT342E, MUH211, MUH212, MUH311, MUH312, MUH321, MAT492”dir.

6.3.1 Veri tablosunun hazırlanması

Bu bölümde modelde kullanılan veri tablosunun hazırlanması anlatılmaktadır.

ODM’de birliktelik kuralları modeli oluşturmak için kullanılan veri “transactional” (işlemsel) olmalıdır. Modelin oluşturulmasında kullanılan işlemsel veri tablosunda bir işlem için birden fazla satırda kayıt bulunmaktadır [11].

Bölüm 5’te anlatılan düzenlemelerin uygulandığı veriler üzerinde Ekler bölümündeki fonksiyonlar çalıştırılarak öğrencilerin MAT ve MUH kodlu derslerden ilk aldıkları harf notu bilgileri çekilmiştir. Her öğrenci için her bir dersi ve dersin çekilen harf notu tek bir satıra kaydedilerek model için veri tablosu oluşturulmuştur.

Oluşturulan tablonun sütun açıklamaları Tablo 6.7’de verilmiştir.

Tablo 6.7: Bölüm dersleri tablosunun sütun açıklamaları

SÜTUN ADI	SÜTUN AÇIKLAMASI	SÜTUN DEĞERLERİ
Öğrenci No	Öğrencinin ID’sini belirtir.	8-497 arası değer içerir.
Ders	Dersin kodunu belirtir.	“MAT111”, “MAT112”, “MAT141”, “MAT142”, “MAT175”, “MAT213”, “MAT214”, “MAT221”, “MAT232”, “MAT242”, “MAT322E”, “MAT331”, “MAT332”, “MAT355”: Ders Planı’ne göre ders alan öğrenci için yoktur, “MAT341E”, “MAT342E”, “MUH211”, “MUH212”, “MUH311”, “MUH312”, “MUH321”, “MAT492”,
Harf_Notu	Dersten alınan ilk harf notunu belirtir.	“AA”, “BA”, “BB”, “CB”, “CC”, “DC”, “DD”, “FF”, “VF”

Tablo 6.7’deki bilgilerle oluşturulan excel belgesi Bölüm 6.2.1.2’de anlatıldığı gibi Oracle Veri Tabanı’na aktarılmış, “BOLUM_DERSLERI” tablosu oluşturulmuştur (Şekil 6.34).

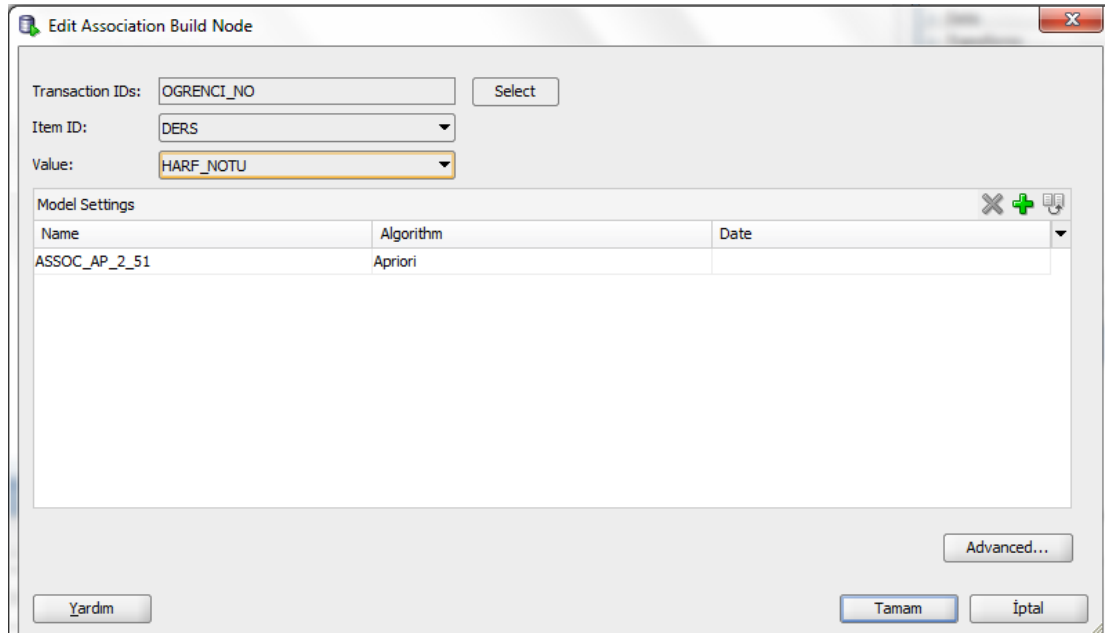
	OGRENCI_NO	DERS	HARF_NOTU
13		8 MAT332	BA
14		8 MAT341E	AA
15		8 MAT342E	AA
16		8 MAT492	AA
17		8 MUH211	AA
18		8 MUH212	AA
19		8 MUH311	BA
20		8 MUH312	BA
21		8 MUH321	BA
22		9 MAT111	BB
23		9 MAT112	BB
24		9 MAT141	BA
25		9 MAT142	AA
26		9 MAT175	AA
27		9 MAT213	AA
28		9 MAT214	AA
29		9 MAT221	AA
30		9 MAT222	BA
31		9 MAT232	BA
32		9 MAT322E	BA
33		9 MAT331	FF
34		9 MAT332	AA
35		9 MAT341E	AA
36		9 MAT342E	AA

Şekil 6.34: “BOLUM_DERSLERI” tablosundan bir görüntü

6.3.2 Model oluşturma

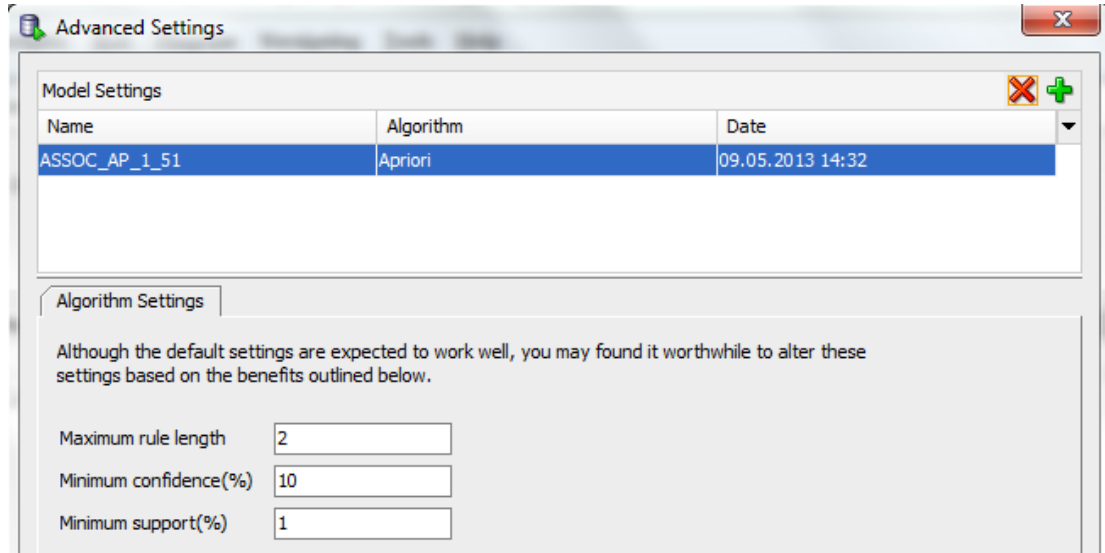
İş akış diyagramına “Data Source” düğümü eklenir ve “BOLUM_DERSLERI” tablo bilgileri yüklenir. İş akış diyagramına “Association” düğümü eklenir ve “Data Source” düğümü ile bağlantısı sağlanır.

“Association” düğümü üzerine gelinerek birliktelik kurallarını çıkaracak model için ayarlar yapılır. Her bir öğrenci bir işleme denk gelir bu sebeple “Transaction IDs” seçeneği; “OGRENCI_NO”, kurallar derslere göre çıkarıldığından “ITEM_ID” seçeneği; “DERS”, dersler arasında çıkarılacak kurallar derslerin harf notlarını baz aldığından “Value” seçeneği; “HARF_NOTU” olarak ayarlanır (Şekil 6.35).



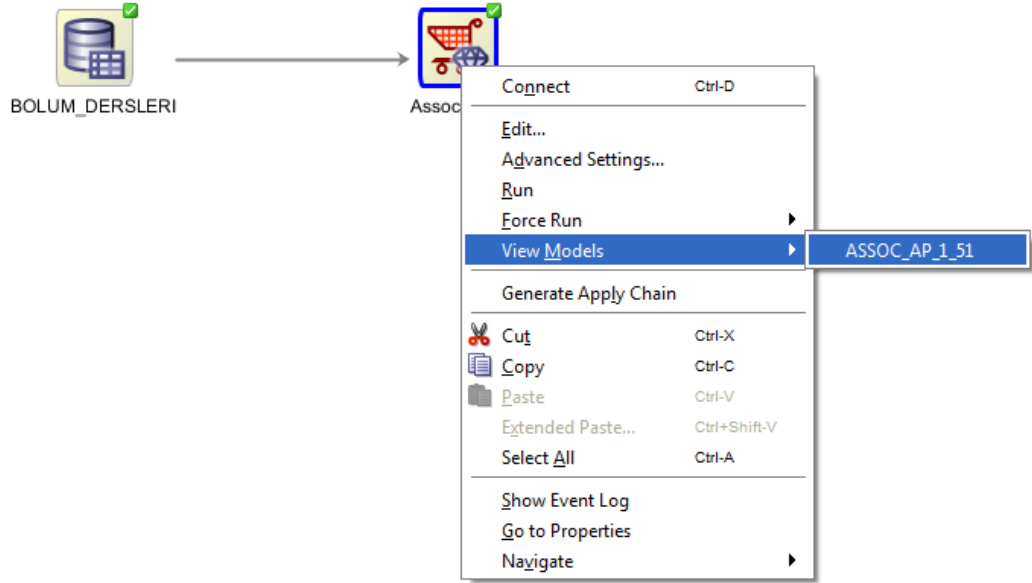
Şekil 6.35: Birliktelik Kuralları model ayarları

“Advanced Settings”ten algoritma ile ilgili destek ve güven değerleri, kuralların en fazla uzunluğu ayarları yapılır. Destek değeri “%1”de ve güven değeri “%10”da varsayılan değerlerde bırakılmıştır. En fazla kural uzunluğu ise “2” olarak ayarlanmıştır (Şekil 6.36).



Şekil 6.36: Algoritma ayarları

Yapılan ayarların ardından “Association” düğümü çalıştırılmıştır. Çalıştırılmış düğüm üzerinde “View Models” seçilerek oluşan birliktelik kuralları görülür (Şekil 6.37).



Şekil 6.37: Model görüntüleme

6.3.3 Birliktelik kuralları

Oluşan birliktelik kuralları “ASSOC_A1_51”de “Rules” sekmesi altından görülür (Şekil 6.38).

ID	Antecedent	Consequent	Lift	Confidence(%)	Support(%)	Item Count	Antecedent Support(%)	Consequent Support(%)
1303	MAT142 = VF	MAT111 = FF	16,1765	100	1,4545	1	6,1818	1,4545
1302	MAT111 = FF	MAT142 = VF	16,1765	23,5294	1,4545	1	1,4545	6,1818
6328	MAT322E = AA	MAT214 = AA	15,7143	57,1429	1,4545	1	3,6364	2,5455
6327	MAT214 = AA	MAT322E = AA	15,7143	40	1,4545	1	2,5455	3,6364
3139	MAT214 = AA	MAT142 = AA	15,2778	50	1,8182	1	3,2727	3,6364
3138	MAT142 = AA	MAT214 = AA	15,2778	55,5556	1,8182	1	3,6364	3,2727
11027	MAT322E = AA	MAT331 = AA	13,0952	57,1429	1,4545	1	4,3636	2,5455
11026	MAT331 = AA	MAT322E = AA	13,0952	33,3333	1,4545	1	2,5455	4,3636
3144	MAT142 = AA	MAT331 = AA	12,7315	55,5556	1,8182	1	4,3636	3,2727
3145	MAT331 = AA	MAT142 = AA	12,7315	41,6667	1,8182	1	3,2727	4,3636

Şekil 6.38: Oluşan kurallardan bir görüntü

“More” seçilerek çıkan “Use Filter” seçeneği işaretlenerek belirlenen kriterlere göre oluşan kurallar üzerinde eleme yapılabilmektedir. Çıkan kuralların en az %70 güvenilirlikte olması istendiğinden “Minimum Confidence(%)”; “70” girilmiştir (Şekil 6.39).

☒ Use Filter

Filter

Minimum Lift:

Minimum Support(%):

Minimum Confidence(%):

Maximum Items In Rule:

Minimum Items In Rule:

Şekil 6.39: Filtreleme ayarları

Bu ayardan sonra yapılan sorguda %70 güvenirliliğin üzerinde 27 kural çıktığı görülmüştür (Şekil 6.40).

Rule Content:

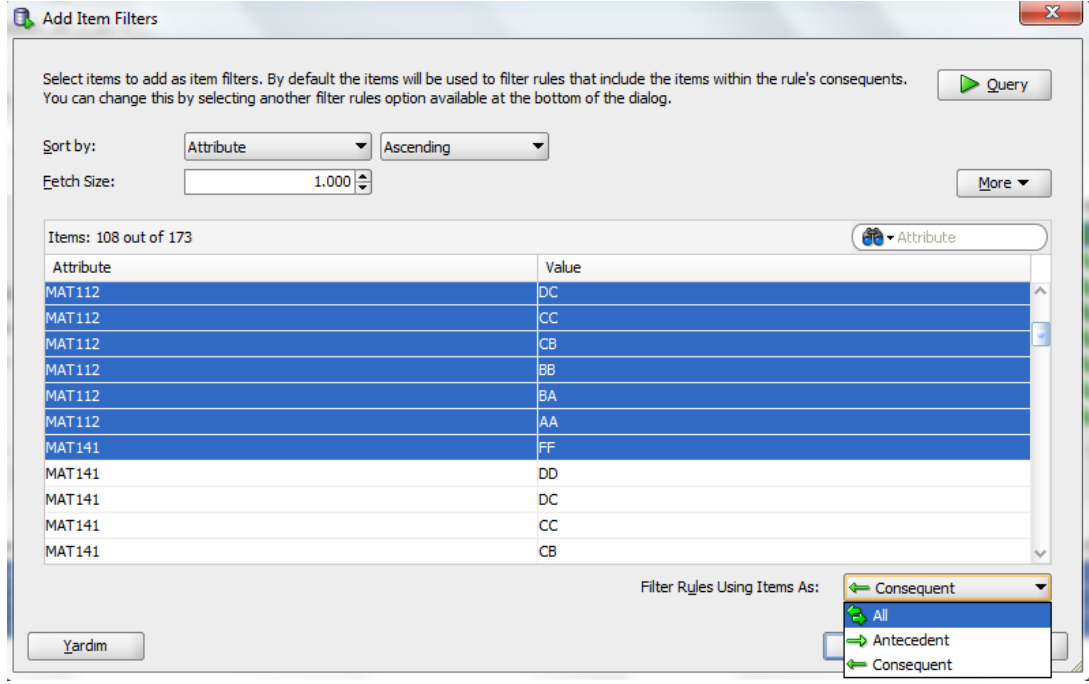
Rules: 27 out of 13.711

ID	Antecedent	Consequent	Lift	Confidence(%)	Support(%)
1303	MAT142 = VF	MAT111 = FF	16,1765	100	1,4545
10053	MUH312 = AA	MAT332 = AA	9,8214	75	1,0909
13644	MAT322E = VF	MAT342E = FF	8,9674	75	1,0909
1718	MAT322E = VF	MAT112 = FF	7,9327	75	1,0909
1622	MAT142 = VF	MAT112 = FF	7,9327	75	1,0909
13519	MAT322E = AA	MAT341E = AA	7,8571	85,7143	2,1818
6331	MAT214 = AA	MAT341E = AA	7,3333	80	2,9091
12207	MAT322E = AA	MUH312 = BB	6,3364	71,4286	1,8182
11034	MAT331 = AA	MAT342E = AA	6,3021	91,6667	4
2192	MAT322E = AA	MAT112 = AA	6,1384	71,4286	1,8182

Şekil 6.40: %70 güvenirliliğin üzerindeki kurallar

Bu kurallar içinde “Ancedent” yani öncül elemanın üçüncü sınıf derslerinden olup, “Consequent” yani sonuç elemanın birinci sınıf derslerinden olduğu kuralların çıktığı görülmüştür. Bu sebeple birinci sınıf dersleri kuralları, ikinci sınıf dersleri kuralları, üçüncü sınıf dersleri kuralları ayrı ayrı çıkarılmıştır.

“Item Filters”dan “Ancedent” ve “Consequent” elemanları belirlenmektedir. Birinci sınıf kuralları için “Item Filters”da “MAT111”, “MAT112”, “MAT141”, “MAT142”, “MAT175” özellikleri için tüm değerler seçilerek “Filter Rules Using Items As” seçeneği “All” seçilmiştir (Şekil 6.41). Böylece bu özellikler hem öncül eleman hem de sonuç elemanı olarak belirlenmiş olur.



Şekil 6.41: Birinci sınıf kuralları için “Item Filters” ayarları

“Item Filters”da yapılan düzenlemenin ardından özellikler Şekil 6.42’deki gibi görülür.

Attribute	Value	Filter
MAT111	FF	Both
MAT111	DD	Both
MAT111	DC	Both
MAT111	CC	Both
MAT111	CB	Both
MAT111	BB	Both
MAT111	BA	Both
MAT111	AA	Both

Şekil 6.42: Birinci sınıf dersleri için özellik ayarları

Oluşturulmuş kurallar üzerinde, belirlenen kriterler ile sorgu çalıştırılır elde edilen sonuç Şekil 6. 43’te görülmektedir. Birinci sınıf dersleri üzerine çıkan kurallardan Bölüm 6.3.4’te bahsedilmektedir.

ID	Antecedent	Consequent	Lift	Confidence(%)	Support(%)	Item Count	Antecedent Support(%)	Consequent Support(%)
1303	MAT142 = VF	MAT111 = FF	16,1765	100	1,4545	1	6,1818	1,4545
1622	MAT142 = VF	MAT112 = FF	7,9327	75	1,0909	1	9,4545	1,4545

Şekil 6.43: Birinci sınıf dersleri kuralları

Aynı şekilde ikinci sınıf kuralları için “Item Filters”da sırasıyla ikinci sınıf dersleri MAT213, MAT214, MAT221, MAT232, MAT242, MAT322E, MUH211, MUH212

derslerinin tüm değerleri hem öncül eleman hem de sonuç elemanı olarak belirlenmiştir. Yapılan ayarlara göre birliktelik kuralları üzerinde çalışan sorgunun çıktısı Şekil 6.44'teki gibidir. İkinci sınıf dersleri üzerine çıkan kurallardan Bölüm 6.3.4'te bahsedilmektedir.

Rules: 3 out of 13.711

ID	Antecedent	Consequent	Lift	Confidence(%)	Support(%)	Item Count	Antecedent Support(%)	Consequent Support(...)
5345	MAT322E = AA	MAT213 = BA	5,4563	71,4286	1,8182	1	13,0909	2,5455
5517	MAT214 = AA	MAT213 = AA	4,2778	70	2,5455	1	16,3636	3,6364
9322	MAT322E = VF	MAT221 = FF	2,7228	100	1,4545	1	36,7273	1,4545

Şekil 6.44: İkinci sınıf dersleri kuralları

Üçüncü sınıf kuralları için de aynı işlem tekrarlanmıştır. MAT331, MAT332, MAT341E, MAT342E, MAT355, MUH311, MUH312, MUH321 üçüncü sınıf derslerinin tüm değerleri, hem öncül eleman hem de sonuç elemanı olarak belirlenmiştir. Yapılan ayarlara göre birliktelik kuralları üzerinde çalışan sorgunun çıktısı Şekil 6.45'teki gibidir. Üçüncü sınıf dersleri üzerine çıkan kurallardan Bölüm 6.3.4'te bahsedilmektedir.

Rules: 3 out of 13.711

ID	Antecedent	Consequent	Lift	Confidence(%)	Support(%)	Item Count	Antecedent Support(%)	Consequent Support(...)
10053	MUH312 = AA	MAT332 = AA	9,8214	75	1,0909	1	7,6364	1,4545
11034	MAT331 = AA	MAT342E = AA	6,3021	91,6667	4	1	14,5455	4,3636
11024	MAT331 = AA	MUH321 = AA	4,2014	91,6667	4	1	21,8182	4,3636

Şekil 6.45: Üçüncü sınıf dersleri kuralları

Sınıflar üzerine çıkarılan kurallar dışında tüm sınıflara ait dersler üzerine çıkan kurallardan sınıf önceliklerine göre anlamlı olan kuralları belirlemek için “Item Filters”da “MAT111”, ”MAT112”, “MAT141”, “MAT142”, ”MAT175”, ”MAT221” özellikleri “Antecent”, “MAT213”, “MAT214”, “MAT232”, “MAT242”, “MAT322E”, “MUH211” özellikleri “All”, “MAT331”, “MAT332”, “MAT341E”, “MAT342E”, “MAT355”, “MAT492”, “MUH212”, “MUH311”, “MUH312”, “MUH321” özellikleri ise “Consequent” olarak belirlenmiştir. Yapılan ayarlara göre birliktelik kuralları üzerinde çalışan sorgunun çıktısı Şekil 6.46'daki gibidir. Çıkan kurallardan Bölüm 6.3.4'te bahsedilmektedir.

Rules: 14 out of 13.711						
ID	Antecedent	Consequent	Lift	Confidence(%)	Support(%)	Item Count
13644	MAT322E = VF	MAT342E = FF	8,9674	75	1,0909	1
13519	MAT322E = AA	MAT341E = AA	7,8571	85,7143	2,1818	1
6331	MAT214 = AA	MAT341E = AA	7,8333	80	2,9091	1
12207	MAT322E = AA	MUH312 = BB	6,3364	71,4286	1,8182	1
5345	MAT322E = AA	MAT213 = BA	5,4563	71,4286	1,8182	1
6333	MAT214 = AA	MAT342E = AA	4,8125	70	2,5455	1
5517	MAT214 = AA	MAT213 = AA	4,2778	70	2,5455	1
5642	MAT142 = VF	MUH311 = FF	3,4958	75	1,0909	1
5643	MAT142 = VF	MUH312 = FF	3,2738	75	1,0909	1
8904	MAT232 = AA	MUH321 = AA	3,2083	70	5,0909	1
5640	MAT142 = VF	MAT331 = FF	3,1250	75	1,0909	1

Şekil 6.46: Tüm sınıflara ait dersler için birliktelik kuralları

6.3.4 Sonuçlar ve tartışma

Birinci sınıf dersleri için çıkan birliktelik kuralları Tablo 6.8’de verilmiştir.

Tablo 6.8: Birinci sınıf birliktelik kuralları

Öncül Eleman	Sonuç Elemanı	Güvenirlilik
MAT142 = VF	MAT111 = FF	100
MAT142 = VF	MAT112 = FF	75

İkinci sınıf dersleri için çıkan birliktelik kuralları Tablo 6.9’da verilmiştir.

Tablo 6.9: İkinci sınıf birliktelik kuralları

Öncül Eleman	Sonuç Elemanı	Güvenirlilik
MAT322E = AA	MAT213 = BA	71,4286
MAT322E = VF	MAT221 = FF	100

Üçüncü sınıf dersleri için çıkan birliktelik kuralları Tablo 6.10’da verilmiştir.

Tablo 6.10: Üçüncü sınıf birliktelik kuralları

Öncül Eleman	Sonuç Elemanı	Güvenirlilik
MUH312 = AA	MAT332 = AA	75
MAT331 = AA	MAT342E = AA	91,6667
MAT331 = AA	MUH321 = AA	91,6667

Tüm sınıf dersleri için çıkan birliktelik kuralları Tablo 6.11’de verilmiştir.

Tablo 6.11: Tüm sınıflar için birliktelik kuralları

Öncül Eleman	Sonuç Elemanı	Güvenirlilik
MAT322E = VF	MAT342E = FF	75
MAT322E = AA	MAT341E = AA	85,7143
MAT214 = AA	MAT341E = AA	80
MAT322E = AA	MUH312 = BB	71,4286
MAT322E = AA	MAT213 = BA	71,4286
MAT214 = AA	MAT342E = AA	70
MAT214 = AA	MAT213 = AA	70
MAT142 = VF	MUH311 = FF	75
MAT142 = VF	MUH312 = FF	75
MAT232 = AA	MUH321 = AA	70
MAT142 = VF	MAT331 = FF	75
MAT322E = VF	MAT331 = FF	75
MAT111 = FF	MAT222 = FF	76,4706
MAT142 = VF	MAT222 = FF	75

KAYNAKLAR

- [1] ITU, 2011. İstanbul Teknik Üniversitesi Lisans eğitim ve Öğretim Yönetmeliği.
- [2] ITU, İstanbul Teknik Üniversitesi Lisans eğitim ve Öğretim Ders Planları, <http://www.sis.itu.edu.tr/tr/dersplan/plan/MAT/>, Aralık 2012
- [3] Yapıcı A., Özel A., Ayça C., 2010. *Oracle Data Miner ile Kredi Ödemeleri Üzerine Bir Veri Madenciliği Uygulaması*, Bitirme Ödevi, ITU.
- [4] Özkan, Y., 2008. *Veri Madenciliği Yöntemleri*, Papatya Yayıncılık Eğitim.
- [5] Kırış A., Demiryürek U., 2005. *Malzeme Talep ve Sarf Miktarlarının Veri Madenciliği Yöntemleri ile Öngörülmesi*.
- [6] <http://www.oracle.com/webfolder/technetwork/tutorials/obe/db/11g/r2/prod/bidw/datamining/ODM11gR2.htm#t2>, Kasım 2012.
- [7] Khair M., Zakhem W., Moucary C., 2012. *Solving Probation and Change of Major Issues in Higher Education Using Data Mining Techniques*, Int. Journal of Multidisciplinary Sciences and Engineering, **3**, 6-11.
- [8] Roulkin D., 2011. *Predicting 6-Year Graduation and High-Achieving and At-Risk Students*, Institutional Research, Assessment & Planning, 1-7.
- [9] http://www.oracle.com/webfolder/technetwork/tutorials/obe/db/sqldev/r30/SQLdev3.0_Import_Export/sqldev3.0_import_export.htm, Kasım 2012.
- [10] <http://www.oralytics.com/2012/02/odm-11gr2attribute-importance.htm>, Şubat 2013.
- [11] http://docs.oracle.com/cd/E11882_01/datamine.112/e12218.pdf, Mart 2013.

EKLER

EK-1: OgrenciAl01 Fonksiyonu

Amaç: Bir ID'ye ait birden fazla satırda bulunan bilgileri tek bir liste elemanında birleştirmek

Girdi: Liste tipinde 1 argüman, listenin yapısı Tablo'da görülmektedir

Tablo: OgrenciAl01 fonksiyonunun liste tipi argümanının yapısı

8	200510	3.07	3.50
8	200520	3.50	3.50
8	200610	3.90	3.55
8	200620	3.50	3.54
9	200320	3.83	3.83
9	200410	3.53	3.67
9	200420	3.31	3.55
9	200510	3.72	3.59

Çıktı:

{ ID, { { satır 1 }, { satır 2 }, ... , { satır n } } }

EK-2: OgrenciAl02 Fonksiyonu

Amaç: Bir ID'ye ait iki ayrı listede bulunan bilgileri tek bir liste elemanında birleştirmek

Girdi: Liste tipinde 2 argüman

Çıktı:

{ ID, { { liste 1 eleman 1 }, { liste 1 eleman 2 }, ... , { liste 1 eleman n } },
{ { liste 2 eleman 1 }, { liste 2 eleman 2 }, ... , { liste 2 eleman n } } }

EK-3: SerCrs Fonksiyonu

Amaç: Bir öğrencinin hangi dersi kaç defa hangi dönemlerde aldığını bulmak

Girdi: 2 liste elemanı (kurs listesi, öğrencinin aldığı kurs listesi) ve 1 kontrol (“0”: seçmeli “1”: zorunlu) elemanı olmak üzere 3 argüman

Çıktı:

{ {Ders İlk, Dersin Kredisi}, {Tekrar Sayısı}, {{Harf Notu İlk, {Dönem İlk, CRN, Notu, Tekrar Durumu}}, { Harf Notu Son, {Dönem Son, CRN, Notu, Tekrar Durumu} } } }

EK-4: OgrenciAl03 Fonksiyonu

Amaç: Bir ID’ye ait sınıfı öncesinde bitirmesi gereken dönem bilgisini bulmak, ders listelerinden almadığı dersleri yazdırmak

Girdi: Liste tipinde 1 argüman

Çıktı: Bir kaydı 6 boyutlu olan liste elemanı

Eleman 1: ID

Eleman 2: { ÖSYM Başarı Sırası, ÖSYM Giriş Türü, ÖSYM Giriş Türü İşareti}

Eleman 3: { { 1. Dönem, 1. Dönem GPA, Toplam GPA }, ... , { n. Dönem, n. Dönem GPA, Toplam GPA } }

Eleman 4: { { 1. Dönem, Akademik Durum}, ... , { n. Dönem, Akademik Durum} }

Eleman 5: { öğrenime başlama dönemi, {1. Sınıfı öncesinde bitirmesi gereken dönem, 2. Sınıfı öncesinde bitirmesi gereken dönem , 3. Sınıfı öncesinde bitirmesi gereken dönem, 4. Sınıfı öncesinde bitirmesi gereken dönem } }

Eleman 6: { { Ders 1, Dersin Kredisi}, {Tekrar Sayısı}, { { Harf Notu 1, { Dönem, CRN, Notu, Tekrar Durumu } }, ... , { Harf Notu Son , { Dönem, CRN, Notu, Tekrar Durumu} } } }

EK-5: Algebra01 Fonksiyonu

Amaç: MAT322E dersini öğretimine 200510 döneminden önce başlayanlar için bir yıl öne çekmek

Girdi: 1 argümanlı liste elemanı

Çıktı: OgrenciAl03 fonksiyonu çıktısı ile aynı yapıda liste elemanı

EK-6: SerCLS Fonksiyonu

Amaç: Öğrencinin aldığı derslerin başlangıç döneminden bitiş dönemine kadar olan tekrar sayılarını ve ilk aldığı harf notu, son aldığı harf notu bilgilerini bulmak. Eğer başlangıç döneminden bitiş dönemine kadar bu dersler arasından almadığı var ise “YYY” ile işaretlemek, hiç almadığı dersler var ise “NNN” ile işaretlemek

Girdi: 3 liste elemanlı argüman (öğrenci ders listesi, başlangıç dönemi, bitiş dönemi)

Çıktı: Her bir ders için boyutu 3 olan liste elemanı.

Eleman 1: { Ders 1, Dersin Kredisi }

Eleman 2: { Tekrar Sayısı, { {Harf Notu İlk, Notu, Puanı}, {Harf Notu Son, Notu, Puanı} } }

Eleman 3: { {Harf Notu, {Dönem 1, CRN, Notu, Tekrar Durumu} } }

EK-7: Sinif Fonksiyonu

Amaç: Öğrencilerin ders bilgilerini sınıflara göre ayırmak

Girdi: 1 liste elemanı (öğrenci ders listesi) ve sınıf derecesi olmak üzere 2 argüman

Çıktı:

Bir öğrenci için 7 boyutlu olan liste elemanı.

Eleman 1: ID

Eleman 2: { ÖSYM Başarı Sırası, ÖSYM Giriş Türü, ÖSYM Giriş Türü İşareti }

Eleman 3: { { 1. Dönem, 1. Dönem GPA, Toplam GPA }, ... , { n. Dönem, n. Dönem GPA, Toplam GPA } }

Eleman 4: { { 1. Dönem, Akademik Durum }, ... , { n. Dönem, Akademik Durum } }

Eleman 5: { öğrenime başlama dönemi, {1. Sınıfı öncesinde bitirmesi gereken dönem, 2. Sınıfı öncesinde bitirmesi gereken dönem , 3. Sınıfı öncesinde bitirmesi gereken dönem, 4. Sınıfı öncesinde bitirmesi gereken dönem } }

Eleman 6: (sınıf ayırt etmeksizin tüm dersler için) { { Ders 1, Dersin Kredisi}, {Tekrar Sayısı}, { { Harf Notu 1, { Dönem, CRN, Notu, Tekrar Durumu } }, ... , { Harf Notu Son , { Dönem, CRN, Notu, Tekrar Durumu} } } }

Eleman 7: Argüman olarak verilen sınıfın dönemlerinde alınan ve verilen sınıfa ait derslerin bilgileri bu elemana SerCLS fonksiyonun çıktısı formatında yerleşmiştir. Bu eleman 3 boyutludur.

Eleman 7.1: verilen sınıfa ait dersler

Eleman 7.2: verilen sınıftan önceki sınıflara ait dersler (verilen sınıftan önce aldığı dersler de bulunur)

Eleman 7.3: verilen sınıftan sonraki sınıflara ait dersler (verilen sınıftan önce aldığı dersler de bulunur)

EK-8: IncDec Fonksiyonu

Amaç: Tekrar aldığı ve ilk aldığı dersleri ortalamasını olumlu ve olumsuz etkileyenler şeklinde ayırmak

Girdi: kaçınıcı sınıf olduğunu belirten kontrol değişkeni, kontrol değişkeninde verilen sınıfın başladığı dönem, ders bilgilerinin bulunduğu liste elemanı, kontrol değişkeninde verilen sınıfa başlarkenki ortalaması (1. Sınıf için bu kontrolü program kullanmamaktadır)

Çıktı: Bir ders için boyutu 3 olan liste.

Eleman 1: { Ders 1, Dersin Kredisi }

Eleman 2: { Tekrar Sayısı, { { ILK, ilk alıp artırdığı tekrar sayısı ("0" veya "1") * kredi, ilk alıp azalttığı tekrar sayısı ("0" veya "1") * kredi, artırdığı not * kredi, azalttığı not * kredi }, { TEK, artırdığı tekrar sayısı * kredi, azalttığı tekrar sayısı * kredi, artırdığı not * kredi, azalttığı not * kredi, { 1. artım, 2. artım, n. artım }, { 1. azalım, 2. azalım, n. azalım } } } }

Eleman 3: { {Harf Notu, {Dönem 1, CRN, Notu, Tekrar Durumu} } }

EK-9: Tablo Fonksiyonu

Amaç: verilen sınıf için ilk alınan ve tekrar edilen dersleri ortalamaya olumlu olumsuz katkıda bulunanlar olarak belirlemek

Girdi: Sınıf Fonksiyonu'nun çıktısı string tipinde

Çıktı: Bir öğrenci için 8 boyutlu olan liste elemanı.

Eleman 1: ID

Eleman 2: { ÖSYM Başarı Sırası, ÖSYM Giriş Türü, ÖSYM Giriş Türü İşareti }

Eleman 3: { { 1. Dönem, 1. Dönem GPA, Toplam GPA }, ... , { n. Dönem, n. Dönem GPA, Toplam GPA } }

Eleman 4: { { 1. Dönem, Akademik Durum }, ... , { n. Dönem, Akademik Durum } }

Eleman 5: { öğrenime başlama dönemi, {1. Sınıfı öncesinde bitirmesi gereken dönem, 2. Sınıfı öncesinde bitirmesi gereken dönem , 3. Sınıfı öncesinde bitirmesi gereken dönem, 4. Sınıfı öncesinde bitirmesi gereken dönem } }

Eleman 6: (sınıf ayırt etmeksizin tüm dersler için) { { Ders 1, Dersin Kredisi}, {Tekrar Sayısı}, { { Harf Notu 1, { Dönem, CRN, Notu, Tekrar Durumu } }, ... , { Harf Notu Son , { Dönem, CRN, Notu, Tekrar Durumu } } } }

Eleman 7: Argüman olarak verilen sınıfın dönemlerinde alınan ve verilen sınıfa ait derslerin bilgileri bu elemana SerCLS fonksiyonun çıktısı formatında yerleşmiştir. Bu eleman 3 boyutludur.

Eleman 7.1: verilen sınıfa ait dersler

Eleman 7.2: verilen sınıftan önceki sınıflara ait dersler (verilen sınıftan önce aldığı dersler de bulunur)

Eleman 7.3: verilen sınıftan sonraki sınıflara ait dersler (verilen sınıftan önce aldığı dersler de bulunur)

Eleman 8: i=1 to 3, j=1 to ders sayısı olmak üzere

Eleman 8.i.j.1: { Ders 1, Dersin Kredisi }

Eleman 8.i.j.2: { Tekrar Sayısı, { {ilk harf notu, notu, puanı}, {son harf notu, notu, puanı} }, { { ILK, ilk alıp artırdığı tekrar sayısı ("0" veya "1") * kredi, ilk alıp azalttığı tekrar sayısı ("0" veya "1") * kredi, artırdığı not * kredi, azalttığı not * kredi }, { TEK, artırdığı tekrar sayısı * kredi, azalttığı tekrar sayısı * kredi, artırdığı not * kredi, azalttığı not * kredi, { 1. artım, 2. artım, n. artım }, { 1. azalım, 2. azalım, n. azalım } } } }

Eleman 8.i.j.3: { {Harf Notu, {Dönem 1, CRN, Notu, Tekrar Durumu} } }

Eleman 8.1: verilen sınıfa ait dersler

Eleman 8.1.1: verilen sınıf 3. sınıf ise 3. sınıfa ait zorunlu dersler

Eleman 8.1.2: verilen sınıf 3. sınıf ise 3. sınıfa ait seçmeli dersler

Eleman 8.2: verilen sınıftan önceki sınıflara ait dersler (verilen sınıftan önce aldığı dersler de bulunur)

Eleman 8.3: verilen sınıftan sonraki sınıflara ait dersler (verilen sınıftan önce aldığı dersler de bulunur)

EK-10: SonTab Fonksiyonu

Amaç: Verilen sınıf için ortalamaya olumlu olumsuz katkıda bulunanlar olarak belirlenmiş ilk alınan ve tekrar edilen derslerin toplamlarını bulma, BIL101E için ayrıca hesap yapma

Girdi: Tablo Fonksiyonu'nun çıktısı string tipinde 1 argüman

Çıktı: Bir öğrenci için 9 elemanlı olan liste elemanı. İlk 8 eleman Tablo fonksiyonunun çıktısı ile aynı yapıda.

Eleman 9:

Eleman 9.1:

“*****”: BIL101E verilen dönemde alınmış, “-----”: BIL101E verilen dönemde alınmamış

Eleman 9.2: İlk alıp artırdığı derslerin toplamı

{ ilk alıp artırdığı tekrar sayısı * kredi tüm dersler için, artırdığı not * kredi tüm dersler için}

Eleman 9.3: İlk alıp azalttığı derslerin toplamı

{ ilk alıp azalttığı tekrar sayısı * kredi tüm dersler için, azalttığı not * kredi tüm dersler için}

Eleman 9.4: Tekrar alıp artırdığı derslerin toplamı

{ tekrar alıp artırdığı tekrar sayısı * kredi tüm dersler için, artırdığı not * kredi tüm dersler için}

Eleman 9.5: Tekrar alıp azalttığı derslerin toplamı

{ tekrar alıp azalttığı tekrar sayısı * kredi tüm dersler için, azalttığı not * kredi tüm dersler için}

Eleman 9.6: {BIL101E'yi ilk aldığı sayı , BIL101E'yi tekrar aldığı sayı}

Eleman 9.1= “*****” => bu eleman vardır

Eleman 9.1= “-----” => bu eleman yoktur

EK-11: Targ01 Fonksiyonu

Amaç: Öğrenimini dördüncü sınıfı bitirmesi gereken dönem veya öncesinde bitirememiş öğrencilerin belirlenmesi ve öğrenimini dördüncü sınıfı bitirmesi gereken dönem veya öncesinde bitirmiş öğrencilerin mezuniyet not ortalamasının bulunması

Girdi: Yok

Çıktı: 2 boyutlu liste elemanı

Eleman 1:

Boyutu öğrenci sayısına eşittir, her bir öğrenci için istenen zamanda mezun olduysa öğrenci not ortalaması mezuniyet ortalaması olarak alınır, mezun olmadıysa “0” olarak alınır.

{ 1. öğrenci not ortalaması, 2. öğrenci not ortalaması, ... }

Eleman 2:

Boyutu öğrenci not ortalaması bilgisi “0” olmayan öğrenci sayısına eşittir

{ 1. öğrenci not ortalaması, 2. öğrenci not ortalaması, ... }

EK-12: Target Fonksiyonu

Amaç: Öğrenimini dördüncü sınıfı bitirmesi gereken dönem veya öncesinde bitirmiş öğrencilerin mezuniyet not ortalamalarının 3 başarı grubuna kümelenmesi

Girdi: 1 argümanlı liste elemanı

{küme1 alt sınır, küme2 alt sınır, küme3 alt sınır }

Çıktı:

Boyutu öğrenci sayısına eşit olan liste elemanı. Elemanlar mezun değil: “0”, Küme1: “1”, Küme2: “2”, Küme3: “3” şeklindedir

EK-13: EkPro Fonksiyonu

Amaç: Öğrencilerin öğrenimleri boyunca iyi durum ve gözetim durumunda olduğu dönem sayısını ve birinci, ikinci, üçüncü sınıf sonlarındaki ortalamalarının bulunması.

Girdi: 1 argümanlı liste elemanı.

{ ID, { ÖSYM Başarı Sırası, ÖSYM Giriş Türü, ÖSYM Giriş Türü İşareti}, { { 1. Dönem, 1. Dönem GPA, Toplam GPA }, ... , { n. Dönem, n. Dönem GPA, Toplam GPA } }, { { 1. Dönem, Akademik Durum}, ... , { n. Dönem, Akademik Durum} }, { öğrenime başlama dönemi, {1. Sınıfı öncesinde bitirmesi gereken dönem, 2. Sınıfı öncesinde bitirmesi gereken dönem , 3. Sınıfı öncesinde bitirmesi gereken dönem, 4. Sınıfı öncesinde bitirmesi gereken dönem } }

Çıktı:

{ ID, { Öğrenimleri boyunca iyi durumda olduğu dönem sayısı, gözetimde olduğu dönem sayısı},

{ { Birinci sınıfın son dönemi, dönem sonu ortalaması, birinci sınıf sonu toplam ortalama}, { İkinci sınıfın son dönemi, dönem sonu ortalaması, ikinci sınıf sonu toplam ortalama}, { Üçüncü sınıfın son dönemi, dönem sonu ortalaması, üçüncü sınıf sonu toplam ortalama} } }

EK-14: IstDers Fonksiyonu

Amaç: Derslerin tekrar sayıları ve derslerin ortalamaları ile ilgili istatistiksel bilgiyi bulmak

Girdi: Dersin adı 1 argüman

Çıktı:

Derslerle ilgili istatistikleri içeren liste elemanı ve derslerle ilgili istatistikleri gösteren pasta grafiği.

İstatistik Liste Elemanı:

4 boyutlu bir listedir.

Eleman 1:

{ dersin toplam kaç kez alındığı, ders öğrenci başına kaç kez alındı (dersin toplam kaç kez alındığı / öğrenci sayısı), dersin genel ortalaması },

Eleman 2:

{ { dersi sadece 1 kez alanların sayısı, dersin dersi sadece 1 kez alanlara göre ortalaması }, { dersi ilk kez alanların sayısı (zorunlu dersler için öğrenci sayısına eşittir), dersin dersi ilk kez alanlara göre ortalaması } }

Eleman 3:

{ { dersi sadece 2 kez alanların sayısı, dersi sadece 2 kez alanların 2. kez aldıkları notların ortalaması }, { dersi en az 2 kez alanların sayısı, dersi en az 2 kez alanların 2. kez aldıkları notların ortalaması } }

Eleman 4:

{ { dersi sadece 3 kez alanların sayısı, dersi sadece 3 kez alanların 3. kez aldıkları notların ortalaması }, { dersi en az 3 kez alanların sayısı, dersi en az 3 kez alanların 3. kez aldıkları notların ortalaması } }

Pasta Grafiği:

Dersin dersi sadece 1 kez alanlara göre ortalaması, dersi sadece 2 kez alanların 2. kez aldıkları notların ortalaması ve dersi sadece 3 kez alanların 3. kez aldıkları notların ortalaması bilgilerini gösteren 3 dilimli grafik.

EK-15: IstDersDonem Fonksiyonu

Amaç: Derslerin dönemlere göre kaç kişi tarafından alındığını ve o dönemki ortalamasının bulunması

Girdi: Ders adını içeren 1 argüman

Çıktı: Derslerle ilgili istatistikleri içeren liste elemanı ve derslerle ilgili istatistikleri gösteren 2 sütun grafiği

İstatistik Liste Elemanı:

Boyutu hesap yapılan dönem sayısından bir fazladır.

{ { Tüm dönemlerde ilk kez alanların sayısı, Dersin tüm dönemlerde ilk kez alanlara göre ortalaması, Tüm dönemlerde dersi tüm alanların sayısı, Dersin tüm dönemlerde tüm alanlara göre ortalaması },

{ Dönem1, Dönem1’de dersi ilk kez alanların sayısı, Dersin Dönem1’de ilk kez alanlara göre ortalaması, Dönem1’de dersi tüm alanların sayısı, Dersin Dönem1’de tüm alanlara göre ortalaması } }

Sütun Grafiği 1:

Döneme göre Dersi ilk kez alanların sayısını ve dersin dersi ilk kez alanlara göre ortalamasını gösterir. Tüm dönemler için dersin dersi ilk kez alanlara göre ortalamasını eşik olarak gösterir.

Yatay=”Dönem”, Düşey=”Ortalama” bilgisini içerir.

Sütun Grafiği 2:

Döneme göre dersi tüm alanların sayısını ve dersin tüm alanlara göre ortalamasını gösterir. Tüm dönemler için dersin dersi tüm alanlara göre ortalamasını eşik olarak gösterir.

Yatay=”Dönem”, Düşey=”Ortalama” bilgisini içerir.

EK-16: IstGozDonem Fonksiyonu

Amaç: Dönemlere göre akademik durum istatistiklerinin belirlenmesi.

Girdi: Yok

Çıktı: Akademik durum istatistiklerini içeren liste elemanı ve derslerle ilgili istatistikleri gösteren 1 sütun grafiği

İstatistik Liste Elemanı:

{Ortalama bitirme dönem sayısı (yaz dönemi de dahil), tüm öğrencilere göre iyi durumda olunan dönem sayısı, tüm öğrencilere göre gözetim durumunda olunan dönem sayısı, tüm öğrencilere göre toplam dönem sayısı, (iyi durumda olunan dönem sayısı) / (toplam dönem sayısı) , {{dönem, toplam öğrenci sayısı, iyi durumda olan öğrenci sayısı, gözetim durumunda olan öğrenci sayısı} ,...}}

Sütun Grafiği:

Döneme göre toplam öğrenci sayısını, iyi durumda olan öğrenci sayısını, gözetim durumunda olan öğrenci sayısını gösterir. “(Tüm dönemlerde iyi durumda olan öğrenci sayısı) / (toplam dönem sayısı)” bilgisi eşik olarak gösterir.

Yatay="Dönem", Düşey="(tüm öğrenciler için iyi durumda olunan dönem sayısı) / (toplam dönem sayısı)" bilgisini içerir.

EK-17: Irregular Öğrenciler için Regular Olma Sütununu Güncelleyen SQL Komutu

```
UPDATE mezun_ogrenci_regularity
```

```
SET regularity=0
```

```
WHERE (donem=200320 OR donem=200420 OR donem=200520 OR  
donem=200620
```

```
OR donem=200720 OR donem=200820 OR donem=200920 OR donem=201020
```

```
OR donem=201120 OR donem=201220 OR donem=201320)
```

EK-18: Regular Öğrenciler için Regular Olma Sütununu Güncelleyen SQL Komutu

```
UPDATE mezun_ogrenci_regularity
```

```
SET regularity=1
```

```
WHERE (donem=200310 OR donem=200410 OR donem=200510 OR  
donem=200610
```

```
OR donem=200710 OR donem=200810 OR donem=200910 OR donem=201010
```

```
OR donem=201110 OR donem=201210 OR donem=201310)
```