

T.C
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

VERİ MADENCİLİĞİ SINIFLAMA
YÖNTEMLERİNİN BAŞARILARININ; BAĞIMLI
DEĞİŞKEN PREVELANSI, ÖRNEKLEM
BÜYÜKLÜĞÜ VE BAĞIMSIZ DEĞİŞKENLER
ARASI İLİŞKİ YAPISINA GÖRE
KARŞILAŞTIRILMASI

Muhsin Özgür DOLGUN

Biyoistatistik Programı
DOKTORA TEZİ

ANKARA
2014

T.C
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

VERİ MADENCİLİĞİ SINIFLAMA
YÖNTEMLERİNİN BAŞARILARININ; BAĞIMLI
DEĞİŞKEN PREVELANSI, ÖRNEKLEM
BÜYÜKLÜĞÜ VE BAĞIMSIZ DEĞİŞKENLER
ARASI İLİŞKİ YAPISINA GÖRE
KARŞILAŞTIRILMASI

Muhsin Özgür DOLGUN

Biyoistatistik Programı
DOKTORA TEZİ

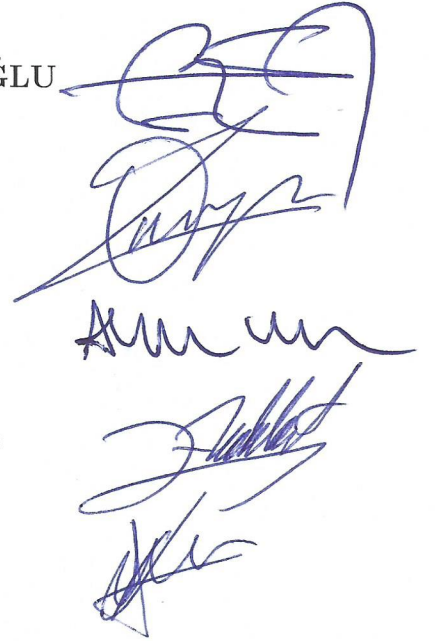
TEZ DANIŞMANI
Prof. Dr. Osman SARAÇBAŞI

ANKARA
2014

Anabilim Dalı : Biyoistatistik
 Program : Biyoistatistik
 Tez Başlığı : Veri Madenciliği Sınıflama Yöntemlerinin Başarı-
 larının; Bağımlı Değişken Prevelansı, Örneklem
 Büyüklüğü ve Bağımsız Değişkenler Arası İlişki
 Yapısına Göre Karşılaştırılması
 Öğrenci Adı-Soyadı : Muhsin Özgür DOLGUN
 Savunma Sınavı Tarihi: 17.09.2014

Bu çalışma jürimiz tarafından Doktora tezi olarak kabul edilmiştir.

Juri Başkanı: Prof. Dr. Ergun KARAAĞAOĞLU
 Hacettepe Üniversitesi
 Tez danışmanı: Prof. Dr. Osman SARAÇBAŞI
 Hacettepe Üniversitesi
 Üye: Prof. Dr. Atilla H. ELHAN
 Ankara Üniversitesi
 Üye: Doç. Dr. Erdem KARABULUT
 Hacettepe Üniversitesi
 Üye: Doç. Dr. S. Kenan KÖSE
 Ankara Üniversitesi



ONAY

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönet-
 meliğinin ilgili maddeleri uyarınca yukarıdaki juri tarafından uygun görülmüş
 ve Sağlık Bilimleri Enstitüsü Yönetim Kurulu kararıyla kabul edilmiştir.


 Prof. Dr. Ersin FADILLIOĞLU
 Müdür Y.

TEŞEKKÜR

Tez çalışmamın gerçekleşmesinde değerli katkı ve yönlendirmelerinden dolayı danışmanım Sayın Prof. Dr. Osman Saraçbaşı'na, değerli görüşleri ile tezin bu noktaya gelmesinde emeği olan tez izleme komitesindeki ve değerlendirme komitesindeki değerli hocalarıma içtenlikle teşekkür ederim.

Son olarak, desteğini her zaman yanımda hissettiğim canım eşim Anıl ve biricik oğlum Derin'e, annem Sıdika ve babam Mehmet'e, bölümdeki meslektaşlarım ve değerli hocalarıma teşekkür ederim.

ÖZET

Dolgun, M. Ö. Veri Madenciliği Sınıflama Yöntemlerinin Başarılarının; Bağımlı Değişken Prevelansı, Örneklem Büyüklüğü ve Bağımsız Değişkenler Arası İlişki Yapısına Göre Karşılaştırılması. Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Programı Doktora Tezi, Ankara, 2014. Karar Ağaçları, Bayesci Ağlar ve Destek Vektör Makinaları literatürde ve uygulamalarda en çok kullanılan istatistiksel ve veri madenciliği kökenli sınıflama yöntemleri arasında yer almaktadır. Bu yöntemler kullanılırken bağımsız değişkenlerin ölçüm seviyesi (sürekli, kesikli vb.), bağımsız değişkenlerin dağılımı (simetrik, çarpık, vb.), bağımlı değişkenin dağılımı (simetrik, çarpık, vb.) bağımsız değişkenler arasındaki ilişki miktarı (düşük, orta veya yüksek ilişki), örneklem hacmi gibi başlıca önemli unsurların model başarısına ve yöntemler üzerine olan etkileri göz ardı edilmektedir. Bu nedenle çalışmada, bağımlı ve bağımsız değişkenlerin farklı yapılarının; Karar Ağaçları, Bayesci Ağlar ve Destek Vektör Makinaları yöntemlerinin başarıları üzerine olan etkileri benzetim çalışması yardımıyla karşılaştırılmıştır. Benzetim çalışmasında, bağımsız değişkenler arasındaki ilişki miktarı için üç farklı seçenek, bağımsız değişken sayısı için üç farklı seçenek, örneklem hacmi için üç farklı seçenek, bağımsız değişkenler ile bağımlı değişken arasındaki ilişki miktarı için üç farklı seçenek, bağımlı değişken prevelansı için de üç farklı seçenek dikkate alınarak toplamda 243 farklı kombinasyon kullanılmıştır. Her bir senaryo 1000 kez tekrar edilmiş, her bir tekrarda sözü edilen sınıflama yöntemleri uygulanmış ve her bir modelin model başarı kriterleri elde edilerek karşılaştırılmıştır. Elde edilen sonuçlar ışığında, araştırmacılara bağımlı ve bağımsız değişkenlerin farklı yapıda olduğu durumlarda hangi sınıflama yöntemini kullanmaları veya hangi yöntemden kaçınmaları konusunda bilgi verilmiştir.

Anahtar Kelimeler: Sınıflama yöntemleri, karar ağaçları, bayesci ağlar, destek vektör makinaları, model karşılaştırma, benzetim çalışması.

ABSTRACT

Dolgun, M. Ö. Comparison of the Performances of Data Mining Classification Methods Based on Prevalence of the Dependent Variable, Sample Size and the Correlation of the Independent Variables. Hacettepe University Institute of Health Sciences, Ph. D. Thesis in Biostatistics, Ankara, 2014. Decision Trees, Bayesian Networks, and Support Vector Machines are the most commonly used statistical and data mining based methods of classification in the literature and practice. While using these methods, the impact of important factors on the model success, such as, the measuring level of the independent variables (i.e., continuous, discrete, etc.), the distribution of the independent variables (i.e., symmetric, skewed, etc.), the amount of correlation between independent variables (i.e., low, medium or strong relationship), and the sample size are often ignored. Therefore, in this study, the impact of different structures of dependent and independent variables on the model performances of Decision Trees, Bayesian Networks, Support Vector Machines methods are compared by a simulation study. A total of 243 different simulation scenarios were obtained by taking into account three levels for the degree of correlation between independent variables, three levels for the number of independent variables in a model, three levels for the sample size, three levels for the amount of the correlation between dependent and independent variables, and three levels for the prevalence of the dependent variable. Each scenario was repeated 1000 times, for each repetition mentioned classification methods are applied and they were compared by their model success criteria. At the end of the thesis, some general suggestions are given to the researchers on which classification method should be used or avoided under different structures of dependent and independent variables.

Key Words: Classification methods, decision trees, bayesian networks, support vector machines, model comparison, simulation study.

İÇİNDEKİLER

	Sayfa
ONAY SAYFASI	iii
TEŞEKKÜR	iv
ÖZET	v
ABSTRACT	vi
İÇİNDEKİLER	vii
SİMGELER VE KISALTMALAR	ix
ŞEKİLLER	x
TABLolar	xiii
1 GİRİŞ	1
2 GENEL BİLGİLER	6
2.1 Veri Madenciliği ve Temel Kavramlar	6
2.1.1 Veri Madenciliği ve Tarihçesi	6
2.1.2 Temel Kavramlar	10
2.1.3 Veri Madenciliği Uygulama Alanları	12
2.1.4 Veri Madenciliğinde Kullanılan Temel Metodoloji: CRISP-DM	14
2.1.5 Veri Madenciliğinde Kullanılan Modeller	18
2.2 Sınıflama Yöntemleri	20
2.2.1 Karar Ağaçları	21
2.2.2 Bayesci Ağlar	25
2.2.3 Destek Vektör Makinaları	27
2.3 Model Başarı Ölçütleri	28
2.3.1 Verinin İki Alt Örnekleme Ayrılması - Hold-Out Yöntemi	29
2.3.2 Genel Doğruluk	29
2.3.3 Dengeli Doğruluk	30
2.3.4 Duyarlılık	30
2.3.5 Seçicilik	30
2.3.6 Matthew Korelasyon Katsayısı	30
2.3.7 F-Ölçütü	31
3 GEREÇ VE YÖNTEM	32

3.1 Benzetim Çalışması	32
4 BULGULAR	37
4.1 Tablo Bazlı Sonuçlar	37
4.2 Grafik Bazlı Sonuçlar	54
4.2.1 C5.0 Yöntemine İlişkin Sonuçlar	55
4.2.2 CHAID Yöntemine İlişkin Sonuçlar	59
4.2.3 C&RT Yöntemine İlişkin Sonuçlar	63
4.2.4 TAN Yöntemine İlişkin Sonuçlar	66
4.2.5 MB Yöntemine İlişkin Sonuçlar	69
4.2.6 DVM Yöntemine İlişkin Sonuçlar	72
5 TARTIŞMA	76
KAYNAKLAR	79
EKLER	
EK 1. BENZETİM ÇALIŞMASI İÇİN KODLAR	
EK 2. MODELLEME ÇALIŞMASI İÇİN KODLAR	

SİMGELELER VE KISALTMALAR

A-CRM	Analitik müşteri ilişkileri yönetimi - Analytic customer relation management
AUC	Eğri altına kalan alan - Area under curve
CHAID	Chi-squared automatic interaction detection
CRISP-DM	Cross industrial processing for data mining
C&RT	Classification and regression tree
DD	Dengeli doğruluk - Balanced accuracy
DVM	Destek vektör makinaları
DWH	Veri ambarı - Datawarehouse
ETL	Veri dönüştürme - Extract transform loading
F	F-ölçütü - F-measure
GD	Genel doğruluk - Overall accuracy
IDC	International data corporation
MB	Markov blanket
MKK	Matthew korelasyon katsayısı - Matthew's correlation coefficient
JDBC	Java database connectivity
ODBC	Open database connectivity
OLAP	Online analytical processing
RBF	Radial Basis Function
DUY	Duyarlılık - Sensitivity
SEC	Seçicilik - Specificity
SQL	Yapısal sorgu dili - Structured query language
TAN	Tree augment naive bayes
TB	Terabytes
VM	Veri madenciliği
VT	Veri tabanı
VTBK	Veri tabanlarında bilgi keşfi - Knowledge discovery in databases
ZB	Zettabytes

ŞEKİLLER

	Sayfa
2.1 Veri Madenciliğinin Tarihsel Gelişim Süreci	9
2.2 Veri Ambarı Mimarisi	11
2.3 CRISP-DM Döngüsü	15
2.4 TAN Örneği	26
2.5 MB Örneği	27
4.1 MKK Kriterine Göre C5.0 Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Örneklem Hacmi Se- naryo Sonuçları	54
4.2 MKK Kriterine Göre C5.0 Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Örneklem Hacmi Se- naryo Sonuçları	55
4.3 MKK Kriterine Göre C5.0 Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Prevelans Senaryo Sonuçları	56
4.4 DD Kriterine Göre C5.0 Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Örneklem Hacmi Se- naryo Sonuçları	57
4.5 DD Kriterine Göre C5.0 Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Prevelans Senaryo Sonuçları	57
4.6 MKK Kriterine Göre CHAID Algoritmasının Bağımsız Korelas- yon × Bağımlı Korelasyon × Değişken Sayısı × Örneklem Hacmi Senaryo Sonuçları	59
4.7 MKK Kriterine Göre CHAID Algoritmasının Bağımsız Korelas- yon × Bağımlı Korelasyon × Değişken Sayısı × Prevelans Senaryo Sonuçları	60
4.8 DD Kriterine Göre CHAID Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Örneklem Hacmi Se- naryo Sonuçları	61
4.9 DD Kriterine Göre CHAID Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Prevelans Senaryo Sonuçları	61
4.10 MKK Kriterine Göre C&RT Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Örneklem Hacmi Se- naryo Sonuçları	63

4.11 MKK Kriterine Göre C&RT Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Prevelans Senaryo Sonuçları	64
4.12 DD Kriterine Göre C&RT Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Örneklem Hacmi Se- naryo Sonuçları	65
4.13 DD Kriterine Göre C&RT Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Prevelans Senaryo Sonuçları	65
4.14 MKK Kriterine Göre TAN Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Örneklem Hacmi Se- naryo Sonuçları	66
4.15 MKK Kriterine Göre TAN Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Prevelans Senaryo Sonuçları	67
4.16 DD Kriterine Göre TAN Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Örneklem Hacmi Se- naryo Sonuçları	68
4.17 DD Kriterine Göre TAN Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Prevelans Senaryo Sonuçları	68
4.18 MKK Kriterine Göre MB Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Örneklem Hacmi Se- naryo Sonuçları	69
4.19 MKK Kriterine Göre MB Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Prevelans Senaryo Sonuçları	70
4.20 DD Kriterine Göre MB Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Örneklem Hacmi Se- naryo Sonuçları	71
4.21 DD Kriterine Göre MB Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Prevelans Senaryo Sonuçları	71
4.22 MKK Kriterine Göre DVM Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Örneklem Hacmi Se- naryo Sonuçları	72
4.23 MKK Kriterine Göre DVM Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Prevelans Senaryo Sonuçları	73
4.24 DD Kriterine Göre DVM Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Örneklem Hacmi Se- naryo Sonuçları	74

4.25 DD Kriterine Göre DVM Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Prevelans Senaryo Sonuçları	74
---	----

TABLOLAR

	Sayfa
2.1 Sınıflama Tablosu	28
3.1 Tez çalışmasında kullanılan benzetim senaryoları	36
4.1 Yöntemlerin Senaryo Bazlı DD Kriterine Göre Sonuçları	38
4.2 Yöntemlerin Senaryo Bazlı MKK Kriterine Göre Sonuçları	40
4.3 Yöntemlerin Senaryo Bazlı F Kriterine Göre Sonuçları	43
4.4 Yöntemlerin Senaryo Bazlı GD Kriterine Göre Sonuçları	46
4.5 Yöntemlerin Senaryo Bazlı DUY Kriterine Göre Sonuçları	48
4.6 Yöntemlerin Senaryo Bazlı SEC Kriterine Göre Sonuçları	51

1 GİRİŞ

Günümüz modern insanının her alışverişinde, her bankacılık işleminde, her hastane başvurusu veya eczane işleminde, her telefon görüşmesinde, sosyal ağlar üzerinde her türlü paylaşım işleminde kaydedilen, uzaktan algılayıcılardan, uydulardan toplanan, devlet ve işletmelerde yapılan işlemler sonucunda saklanan veriler her saniye inanılmaz boyutlarda artmaktadır. Uluslar arası pazar araştırma firmalarından olan IDC'nin 2013 yılının Aralık ayında açıkladığı raporda, sadece 2012 yılında dünya genelinde toplanan ve yeniden düzenlenen veri boyutunun 2.8 zettabayt (ZB) olduğu tahmin edilmektedir. 1 ZB'ın yaklaşık 1 milyar terabayt (TB) olduğu düşünüldüğünde, artışın boyutları daha açık bir şekilde görülebilmektedir. Aynı şirketin 2020 yılındaki tahmini ise 40 ZB'tır. Diğer bir deyişle; her gün onbinlerce megabyte veri dünya çapında üretilmektedir. Ancak bu hızlı gelişimin izlenmesi, bu artışın üstel olarak artmasından dolayı çok mümkün olmamaktadır. Ne yazık ki her geçen gün daha az bilgi taşıyan daha fazla veri üretilmektedir (1).

Son yıllarda bilgi sistemleri ve teknolojinin gelişmesi sonucunda; kamu kurumları, işletmeler ve diğer kuruluşlar veri tabanlarında kuruluşunun amacına ve yapısına bağlı olarak çeşitli türlerde veri toplamıştır. Uygun yazılımların gelişimi ve firmaların topladığı veriyi kullanılabilir bilgiye çevirme isteği toplanan bu veriyi işleyerek, verinin içerisindeki kullanılabilir ve ilginç ilişkilerin, birlikteliklerin ve örüntülerin ortaya çıkarılmasını gerekli hale getirmiştir. Fakat ulaşılmak istenen bu yararlı, kullanışlı bilgiler çoğunlukla saklı ve henüz işlenmemiş durumdadır.

Saklı ve işlenmemiş veriye olan bu büyük ilgi, Veri Madenciliği (VM) gibi yeni alanların keşfiyle anlaşılabilir ve yorumlanabilir bir hale gelmiştir. Özellikle bilgi teknolojilerindeki dikkati çekecek ölçüde yaşanan gelişmeler ile depolanan verilerin boyutlarının artması, VM ve yapay zeka algoritmalarının gelişmesine sebep olmuştur. Bilgisayar teknolojisinin ve bu yeni alanların kullanımı ile madencilikte altın damarının bulunması gibi binlerce veri yığıntısı arasından bilgi damarı otomatik olarak keşfedilmektedir.

Son zamanlarda VM yöntemleri hem tanımlayıcı (descriptive) hem de çıkarımsal (inferential) yöntemler arasında sıklıkla kullanılır olmuş ve popüler hale gelmiştir. Bunun sebeplerinden belki de en önemlisi, bu yöntemleri uygulayabileceğimiz bilgisayar programlarının hızlı bir şekilde gelişmesi, verinin çok hızlı bir biçimde büyümesi ve toplanabilmesidir. Diğer bir sebep ise klasik istatistiksel yöntemlerin varsayımlarının bozulduğu veri kümelerinde özellikle VM öngörüşel yöntemlerinin daha başarılı sonuçlar göstermesidir (2).

VM'de modelleme algoritmaları öğrenme yöntemlerine bağlı olarak; Danışmanlı (Supervised) ve Danışmansız (Unsupervised) öğrenme olmak üzere ikiye ayrılmaktadırlar. Danışmanlı öğrenme yöntemlerinde bir bağımlı değişken varken, danışmansız öğrenme yöntemlerinde bağımlı değişken kavramı yoktur. VM'de modelleme algoritmaları kullanım alanına bağlı olarak ise; Sınıflama ve Regresyon, Kümeleme ve Birliktelik Kuralları olmak üzere üçe ayrılmaktadırlar. Sınıflama ve regresyon, bir bağımlı değişkeni tahmin etmeye yarayan; kümeleme, eldeki veri kümesini homojen gruplara ayırmaya yarayan ve birliktelik kuralları ise birlikte hareket eden davranış kalıplarını bulmaya çalışan algoritmalarlardır.

Sınıflama yöntemleri, gözlemleri veya bireyleri ait oldukları sınıfa veya kategoriye en az hata ile atamaya çalışan modelleme yöntemlerini kapsayan genel bir terimdir. Bu yöntemler; patern tanımlama, ayırma veya danışmanlı öğrenme olarak da literatürde kullanılmaktadır. İstatistik literatüründe sınıflama yapmak için sıklıkla kullanılan yöntemler, doğrusal ayırma analizi ve lojistik regresyon analizidir. İlgili yöntemler için var olan bazı varsayımlar ve bu varsayım bozulmalarından dolayı; Karar Ağaçları, Destek Vektör Makinaları (DVM), Bayesci Ağlar, k-En Yakın Komşular, Yapay Sinir Ağları, Genetik Algoritmalar, Fuzzy Set Yaklaşımı gibi veri madenciliği yöntemlerinden yararlanılmaktadır (3-8).

VM literatüründe sıklıkla kullanılan sınıflama yöntemleri; Karar Ağaçları, DVM ve Bayesci Ağlardır. İlgili yöntemler, danışmanlı öğrenme yöntemleri ve sınıflama ve regresyon algoritmaları arasında yer almaktadır. Bu tez kapsamında da Karar Ağaçları (Chi-squared Automatic Interaction Detection - CHAID, Classification and Regression Trees - C&RT ve C5.0) , DVM (Radial Basis Function - RBF) ve Bayesci Ağlar (Tree Augmented Naive Bayes - TAN ve Markov Blanket - MB) uygulamada ve literatür çalışmalarında en çok kullanılan algoritmalar olmasından dolayı karşılaştırılacak yöntemler olarak ele alınmıştır. Yine hem uygulamalarda hem de literatürde en sık karşılaşılan sınıflama problemi, iki durumlu (Örneğin; iyi - kötü, hasta - sağlıklı, riskli - riskli değil, kanser - kanser değil) sınıflama problemidir. Bu sebepten dolayı, tez kapsamında bağımlı değişken iki durumlu kategorik değişken olarak tanımlanmıştır.

Karar Ağaçları, adından da anlaşılabilceği gibi ağaç görünümünde tahmin edici bir yöntemdir. Karar ağacı yöntemini kullanarak verinin sınıflanması iki basamaklı bir işlemdir. İlk basamak öğrenme basamağıdır. Öğrenme basamağında önceden bilinen bir eğitim verisi, model oluşturmak amacı ile sınıflama algoritması tarafından çözümlenir. Öğrenilen model, sınıflama kuralları veya karar ağacı olarak gösterilir. İkinci basamak ise sınıflama basamağıdır. Sınıflama basamağında test verisi, sınıflama kurallarının veya karar ağacının doğruluğunu belirlemek amacıyla kullanılır. Eğer doğruluk kabul edilebilir oranda ise, kurallar

yeni verilerin sınıflanması amacıyla kullanılır.

Bayesci Ağlar, bir veri kümesindeki değişkenler arasındaki olasılıksal ve koşullu bağımsızlığı gösteren ve değişkenler arasındaki nedensel ilişkileri açıklayan grafiksel bir model oluşturur. Küçük veri kümelerinde başarılı sonuçlar vermesinden ve bağımsız değişkenler arasındaki ilişkileri grafiksel olarak açıklamasından dolayı sıklıkla tercih edilen yöntemler arasındadır.

DVM yöntemi, 1992 yılında Vapnik tarafından geliştirilmiş ve son yıllarda özellikle VM'de doğrusal ve doğrusal olmayan verilerin sınıflandırılmasında kullanılan bir danışmanlı öğrenme yöntemidir. DVM, sınıflama işlemini kullandığı çekirdek fonksiyonlar yardımıyla gerçekleştirmektedir (7).

Literatürde ve uygulamada sınıflama amaçlı en çok kullanılan yöntemler arasında yer alan Karar Ağaçları, Bayesci Ağlar ve DVM gibi yöntemlerde karşılaşılabilecek sorunlar genelde göz ardı edilmektedir. Bu sorunlar ağırlıklı olarak bağımsız değişkenlerin dağılımları (simetrik, çarpık vd.), ölçüm seviyeleri (kategorik, sürekli), bu değişkenler arasındaki ilişki miktarları (düşük, orta, yüksek) ve örneklem hacminden kaynaklanmaktadır (3, 9, 10).

Klasik öngörüşel yöntemlerin sonuçlarını ve model başarısını ciddi anlamda etkileyen, bağımlı değişkenin dağılımı, bağımsız değişkenler arasındaki ilişki miktarı, örneklem hacmi, bağımsız değişken sayısı ve bağımsız değişkenler ile bağımlı değişken arasındaki ilişki miktarı çoğunlukla VM yöntemleri arasında yer alan öngörüşel yöntemleri de etkilemektedir.

Literatürde daha önce yapılan çalışmalarda;

Entezari-Maleki vd.'nin 2009 yılında yapmış oldukları benzetim çalışmasında; Karar Ağaçları, k-En Yakın Komşular, Lojistik Regresyon, Naïve Bayes ve DVM sınıflama yöntemleri eğri altında kalan alan (AUC) değerlendirme kriterine göre karşılaştırmıştır (11).

Tjen-Sien vd.'nin 2000 yılında hem benzetim yöntemi hem de gerçek veri kullanarak yapmış oldukları çalışmada; Karar Ağaçları, Yapay Sinir Ağları ve İstatistiksel sınıflama yöntemleri algoritma çalışma zamanı ve genel doğruluk kriterleri açısından karşılaştırılmıştır (12).

Othman ve Yau'nun 2007 yılında göğüs kanseri verileri kullanılarak yapmış oldukları çalışmada; Bayesci Ağlar, Karar Ağaçları, Yapay Sinir Ağları ve k-En Yakın Komşular sınıflama yöntemleri Kappa test istatistiği yardımıyla karşılaştırılmıştır (5).

Romera vd.'nin 2008 yılında Cordoba Üniversitesi Öğrenci verileri kullanılarak yapmış oldukları çalışmada; k-En Yakın Komşular, Karar Ağaçları, Apriori Algoritması ve Yapay Sinir Ağları sınıflama yöntemleri genel doğruluk kriterine göre karşılaştırmıştır (6).

Huang vd.'nin 2003 yılında gerçek veri kullanarak yapmış oldukları çalışmada; Karar Ağaçları, DVM ve Bayesci Ağları sınıflama yöntemleri GD ve AUC kriterine göre karşılaştırılmıştır (13).

Amor vd.'nin 2004 yılında Network Saldırısı verilerini kullanarak yapmış oldukları çalışmada; Bayesci Ağları ve Karar Ağaçları yöntemleri karşılaştırılmıştır (14).

Bağımsız değişkenler arasında yüksek ilişki (veya düşük ilişki) olduğu durumlarda ve bağımsız değişkenlerin simetrik veya çarpık dağılması durumlarında sınıflama yöntemlerinin performansının nasıl etkilendiğine dair çalışmalar neredeyse hiç incelenmemiş bir konudur. İncelenen çalışmalarda ise model performansların değerlendirmesi genelde tek bir karşılaştırma kriterine göre yapılmış ve sonuçlar genellenmemiştir.

Bu çalışmada bağımlı ve bağımsız değişkenler arasındaki ilişkilerin farklı yapılarda olduğu durumlarda, Karar Ağaçları, Bayesci Ağlar ve DVM yöntemlerinin model başarılarının benzetim çalışması yardımıyla karşılaştırılması **amaçlanmıştır**. Çalışmanın sonucunda araştırmacılara bağımlı ve bağımsız değişkenlerin farklı yapıda olduğu durumlarda hangi sınıflama yöntemini kullanmaları veya hangi yöntemden kaçınmaları konusunda bilgi verilmiştir.

Yukarıda bahsedilen ve literatürde yapılan çalışmaların çoğunda bu tez kapsamında olan ve olmayan sınıflama yöntemleri karşılaştırılmıştır. Ancak bu çalışmalarda belirli problemlere ilişkin veri kümeleri üzerinde çalışma yapılması (kanseri verisi, network saldırısı verisi, öğrenci verileri vd.) ve sonuçların detaylı bir benzetim çalışması ile genellenememesi bu tez kapsamında ele alınan konunun belirlenmesinde en önemli nokta olmuştur (5, 6, 11–14).

Yukarıdaki varsayım bozulmalarına ek olarak örneklem büyüklüğü ve bağımsız değişkenlerin bağımlı değişken ile olan ilişkisinin, sınıflama yöntemlerinin performansına olan etkisinin incelendiği herhangi bir çalışma yoktur.

Küçük örneklem büyüklüklerinde Bayesci Ağların diğer yöntemlere göre daha başarılı sonuç gösterdiği, çok büyük örneklem büyüklüklerinde Karar Ağaçları ve DVM'nin diğer yöntemlerden daha başarılı sonuç gösterdiği, bağımsız değişkenler arasında ilişki yüksek olduğunda Karar Ağacı ve Bayesci Ağların diğer yöntemlere göre daha başarılı sonuç gösterdiği **bu tezin hipotezlerini oluşturmaktadır**.

Bu amaç doğrultusunda hazırlanan tez çalışmasının giriş bölümünde çalışmanın amacı, kapsamı, daha önce yapılan çalışmalar ve veri madenciliği ile ilgili genel bilgiler verilmiştir. İkinci bölümde, VM kavramı ve tarihçesi, temel kavramlar, VM'de kullanılan bir metodoloji olan CRISP-DM ve VM'de kullanılan modelleme yöntemleri hakkında bilgi verilmiştir. Üçüncü bölümde,

sınıflama yöntemleri ve tez kapsamında kullanılan sınıflama yöntemlerinden olan karar ağaçları, bayesci ağlar ve DVM ile ilgili bilgi verilmiştir. Dördüncü bölümde, verinin alt örneklemelere ayrılması mantığı, Matthews korelasyon katsayısı, dengeli doğruluk, F-ölçütü, duyarlılık, seçicilik ve genel doğruluk model karşılaştırma ölçütleri ve tez kapsamında kullanılan karşılaştırma ölçütleri hakkında bilgi verilmiştir. Beşinci bölümde, senaryolar ve yapılan benzetim çalışması ile ilgili bilgi verilmiştir. Altıncı bölümde, tablo ve grafik bazlı sonuçlar verilmiş, yedinci bölümde ise, sonuçlara ve gelecek çalışmalara ilişkin tartışma gerçekleştirilmiştir.

2 GENEL BİLGİLER

2.1 Veri Madenciliği ve Temel Kavramlar

Son yıllarda bilgi sistemleri ve teknolojinin gelişmesi sonucunda; kamu kurum ve kuruluşları, işletmeler ve diğer kuruluşlar veri tabanlarında kuruluşun amacına ve yapısına bağlı olarak çeşitli türlerde veri toplamaktadır. Fakat bu veriler işlenmediği sürece anlamsız bir veri yığını olarak veri tabanlarında depolanmaktadır (14).

Uygun yazılımların gelişimi ve firmaların topladığı veriyi kullanılabilir bilgiye çevirme isteği toplanan bu veriyi işleyerek, verinin içerisindeki kullanılabilir ve ilginç ilişkilerin, birlikteliklerin ve örüntülerin ortaya çıkarılmasını gerekli hale getirmiştir. Günümüzde pek çok kurum verilerini yararlı ve kullanışlı bilgiler elde edecek yöntemler ile işlemeye başlamamıştır. Ham veri zengini, nitelikli bilgi fakiri durumunda olan kurumların rekabetçi piyasada başarılı olmaları ve başarılarını sürdürmeleri her geçen gün daha da zorlaşmaktadır. Veri toplamanın önemini kavramış olan ve geçmişe yönelik sorgularla veriden en üst düzeyde fayda sağlayamayacağını görmeye başlayan bütün kurumların asıl ihtiyacı verilerin doğru yöntemler ile analiz edilebilmesidir (15).

2.1.1 Veri Madenciliği ve Tarihçesi

Genel tanımı ile VM'ni tanımlamak gerekirse; büyük veri tabanlarından önceden bilinmeyen, kullanışlı bilginin elde edilebilmesi için kullanılan istatistiksel, matematiksel ve yapay zeka kökenli yöntemler bütünü olarak tanımlanabilir. Kullanışlı bilgi, veride önceden bilinmeyen veya şüphelenilmeyen örüntü ve ilişkileri içermektedir. VM kavramı zaman zaman veri tabanlarında bilgi keşfi (VTBK) olarak da kullanılmaktadır. Veri madenciliği kısaca, ham veriden nitelikli bilgiye ulaşmak için atılan adımlar olarak tanımlanabilir.

VM ile ilgili literatürde kullanılan bazı tanımlar aşağıda verilmiştir.

- VM, verideki anlamlı, orijinal, potansiyel olarak kullanışlı ve sonunda anlaşılabilir olan örüntülerin tanımlanmasında önemli bir süreçtir.
- VM, önceden bilinmeyen, anlaşılabilir ve etkin bilginin büyük veri tabanlarından çıkarılma sürecidir ve çıkarılan bu bilgi kritik ticari karar vermede kullanılır.
- VM, bilgi keşfi sürecinde veri içerisindeki önceden bilinmeyen ilişkileri ve

örüntüleri ayırt etmek için kullanılan yöntemler kümesidir.

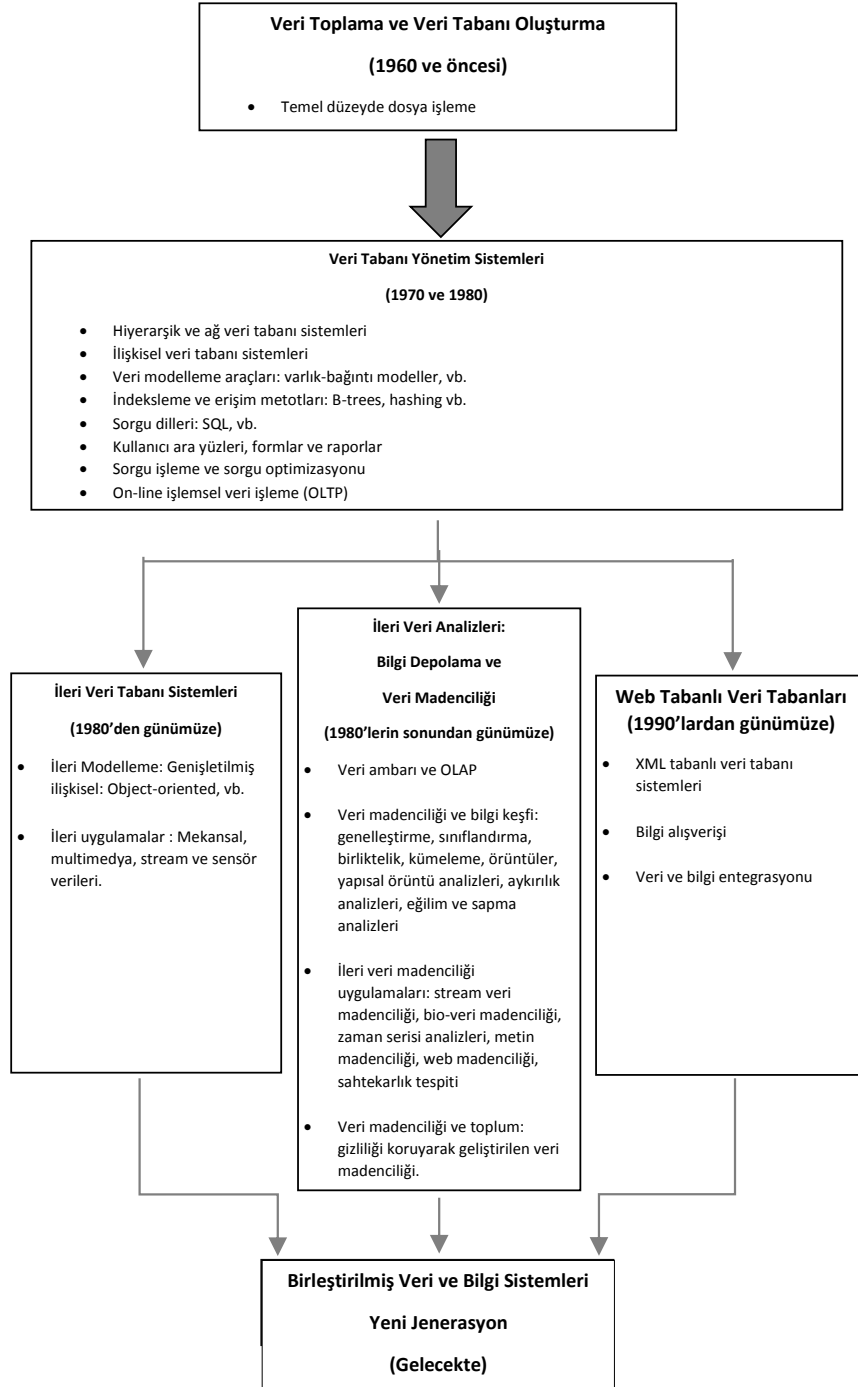
- VM, bilinmeyen ve beklenmeyen örüntülerin büyük veri tabanlarında arandığı bir karar destek sürecidir.
- VM, saklı fakat potansiyel olarak kullanışlı bilgiyi bulmak için kullanılan arama ve veri çözümleme sürecidir.
- Gartner Group tarafından yapılan diğer bir tanımda VM, istatistiksel ve matematiksel yöntemlerle birlikte örüntü tanıma teknolojilerini kullanarak, depolama ortamlarında saklanmış bulunan veri yığınlarının elenmesi ile anlamlı yeni ilişki, örüntü ve eğilimlerin keşfedilmesi süreci olarak ifade edilmektedir (2).

VM ile ilgili diğer tanımlar (4, 8, 10, 16, 17)'da detaylı olarak verilmiştir.

VM, bilişim sürecinin doğal bir sonucu olarak gelişmiş ve bugünkü noktasına gelmiştir. İnsanlığın başlangıcından günümüze veri her zaman biriktirilmiş, yorumlanmış, veriden anlamlı bilgi elde edilmek istenmiş ve bunun için de yazılımlar ve donanımlar geliştirilmiştir. Bu sayede de veri ve bilgi, geçmişten günümüze taşınır hale gelmiştir. 1950'li yıllarda ilk bilgisayarlar basit sayma ve diğer basit aritmetik işlemleri için kullanılmaya başlamıştır. 1960'larda ise veri tabanı ve verilerin depolanması kavramı bilişim dünyasında yerini almıştır. 1960'ların sonunda veri tabanları sadece depolama amacı ile değil aynı zamanda işlem yapma kapasitelerinden dolayı da kullanılmaya başlanmıştır. Bilim insanları basit öğrenme algoritmalarını bilgisayar destekli olarak kullanılabilir şekle sokabilmişlerdir. Minsky ve Papert, günümüzde sinir ağları olarak bilinen algılayıcıların (perceptron) çok basit kuralları öğrenebileceğini göstermişlerdir. 1970'lerde E.F. Codd'un yayınladığı ilişkisel veri tabanları hakkındaki makalesi sonucunda insanlar ilişkisel veri tabanları ile tanışmış oldular. Bilim insanları bununla beraber basit anlamda makine öğrenimini gerçekleştirmişlerdir. Aynı yıllarda, Sistem R, MS SQL Server, Sybase geliştirilmiştir. 1970'li yılların sonunda P. Chen tarafından varlık bağıntı modelleri önerilmiştir. 1980'lerde SQL satandart dil olarak kabul edilmiştir. Veri tabanı yönetim sistemleri yaygınlaşmış, bilimsel ve kamusal alanlarda, özel sektörde sıklıkla kullanılmaya başlanmıştır. Bu yıllarda şirketler, ürünleri, müşterileri ve rakipleri ile ilgili verilerin toplandığı ve bugünkü adıyla A-CRM (analitik müşteri ilişkileri yönetimi) projelerinin temelini oluşturan veri tabanlarını oluşturmuştur. 1990'larda Visual Basic, ODBC, Excel ve Access kavramları bilişim dünyasında yerini almıştır. Artık içindeki veri miktarı katlanarak artan veri tabanlarından, faydalı bilgilerin sadece sorgu ve raporlama sistemleri ile elde edilemeyeceği kabul edilmiş ve nasıl bulunabileceği düşünölmeye

başlanmıştır. Bunun üzerine çalışmalara ve uluslararası çalıştaylara başlanmıştır. 1989 yılında Gregory Piatetsky-Shapiro'nun organize ettiği ilk VTBK çalıştayını gerçekleştirilmiştir. 1995 yılında bu çalıştay, "ACM SIGKDD Conference on Knowledge Discovery and Data Mining" adı altında düzenli hale getirilmiştir. (IJCAI)-89 VTBK Çalışma Grubu toplantısı ve 1991, VTBK (IJCAI)-89'un sonuç bildirgesi sayılabilecek "Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop" makalesinin VTBK ile ilgili temel tanım ve kavramları ortaya koyması ile süreç daha da hızlanmış ve nihayet 1992 yılında veri madenciliği için ilk yazılım gerçekleştirilmiştir. 1990'lı yılların sonu ve 2000'li yıllarda VM sürekli gelişmiş ve hemen hemen tüm alanlara uygulanmaya başlanmıştır. Google, Facebook, Tweeter, LinkedIn gibi popüler ticari şirketlerin ve kamusal güvenlik ile ilgilen kurumların ciddi yatırım yapması ve alana olan ilgiyi daha da artırmıştır (4, 18, 19).

Veri madenciliğinin tarihsel gelişim süreci aşağıdaki Şekil 2.1'de verilmiştir.



Şekil 2.1. Veri Madenciliğinin Tarihsel Gelişim Süreci

Bütün bu sürecin sonunda ortaya çıkan nokta, toplanan verinin doğru yöntemler ile analiz edilmesi ve elde edilen sonuçların sistem içerisinde çalışabilirliğinin sağlanmasıdır.

2.1.2 Temel Kavramlar

VM literatüründe sıklıkla kullanılan bazı temel kavramlar bu bölümde ele alınmıştır. Bu kavramların çoğu bilgisayar mühendisliği literatüründe de karşılaşılan kavramlardır.

Veri madenciliğinin amacı doğrultusunda, genel süreci aşağıdaki gibi özetlemek mümkündür;

$$\text{HAM VERİ} \Rightarrow \text{VERİ} \Rightarrow \text{BİLGİ} \Rightarrow \text{NİTELİKLİ BİLGİ}$$

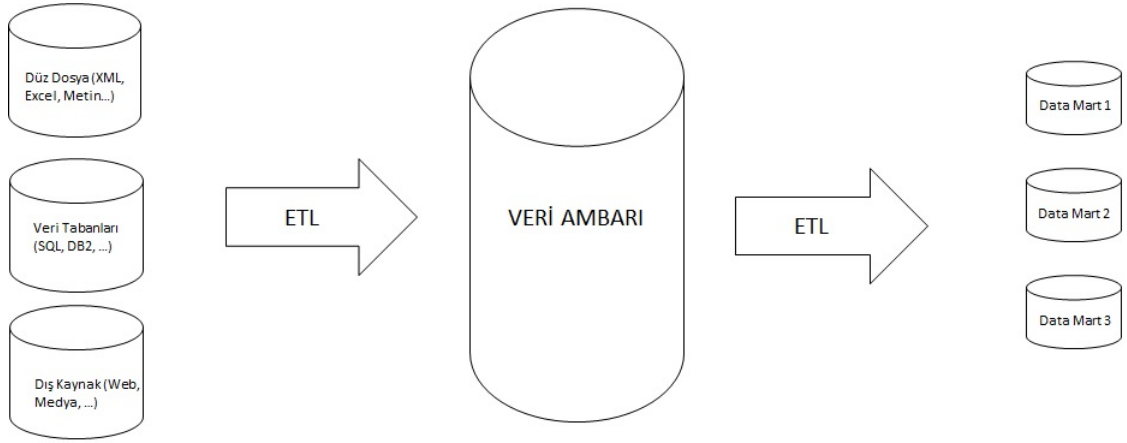
Ham Veri (Raw Data): Düzenlenmemiş veri olarak ifade edilebilir.

Veri (Data): Satır ve kolonlar bazında ifade edilmiş, bir formata sahip olan verileri ifade etmektedir. Veriler ölçüm, sayım, deney, gözlem ya da araştırma yolu ile elde edilmektedir. Ölçüm ya da sayım yolu ile toplanan ve sayısal bir değer bildiren veriler nicel veriler, sayısal bir değer bildirmeyen veriler de nitel veriler olarak adlandırılmaktadır. Kategorik ve sürekli olarak ayrımını yapacağımız bu yapı daha da derin bir bileşen yapısına sahiptir.

Bilgi (Information): Sorgu ve raporlama fonksiyonları sayesinde, veri tabanındaki verinin çekilerek bilgiye dönüşümü sağlanır. Örneğin, ürün, miktar ve fiyat toplamaları, satılan ürünlerle bunların miktar ve hacimleri bilgiyi sağlar.

Nitelikli Bilgi (Knowledge): Genellikle öngörüsül modelleme yöntemleri ile elde edilen sonuç olarak değerlendirilebilir. Sıklıkla elde edilen modelin, tahmin amacıyla kullanılması durumunu ifade etmek için kullanılmaktadır.

Genel yapıya, yine veri madenciliği literatüründe sıkça kullanılan ETL (veri dönüştürme), DWH (veri ambarı), Data Mart gibi terimlerin perspektifinden bakmak yararlı olacaktır. Şekil 2.2 bu yapıyı geneli itibarıyla anlatmaktadır.



Şekil 2.2. Veri Ambarı Mimarisi

Veri Kaynakları: Verinin tutulduğu alanları ifade eder. Veri tabanları, metin dosyaları, excel dosyaları, XML dosyaları, sav ve sas uzantılı dosyalar gibi kaynaklar veri kaynakları olarak nitelendirilirler. Veri kaynakları, özellikle de açık kaynaklı bilgisayar programlarının artmasıyla beraber çok fazla çeşitliliğe sahip olmuştur. Ücretsiz veya ticari veri analiz programlarının tamamı standart olarak kabul edilen veri tabanı, XML ve text veri kaynaklarına bağlanabilmektedir.

Veri Dönüştürme (ETL): ETL kısaltmasında kullanılan harfler, İngilizce karşılığı olan “veriyi çıkar, dönüştür ve yükle” işlemlerini ifade eder.

Projelerde veri her zaman aynı biçimlerde ve istediğiniz detay veya özet durumunda bulunmamaktadır. Zaman zaman ihtiyaca göre aynı verinin kullanım amacına bağlı olarak, farklı düzenlerde kullanılması gerekebilmektedir.

ETL araçları ile verilerinin kaynaklarından ham olarak işlenecek kaynağa çevrilmesi işlemi yapılır. Bu aşamada farklı erişim protokolleri (ODBC, JDBC, doğal) ve dosya biçimleri kullanılır. Elde edilen veriler incelenir, çeşitli filtreleme, temizlik, eşleme, sıralama, ek bilgiler ile zenginleştirme, ayrıştırma gibi işlemler uygulanır.

Veri Ambarı: Veri ambarı, iş hedefleri doğrultusunda sorgulamalar ve analiz yapmak için özelleşmiş bir veri tabanıdır. Temel amacı, işletmeye ait güncel olmayan kayıtları saklamak ve bu kayıtlar üzerinde daha kolay analizler yapılmasını sağlayarak iş ihtiyaçlarını anlamaya ve işletme fonksiyonlarını yenilemeye yardımcı olmak, yani iş zekasına kolaylık sağlamaktır.

“Data Mart”: “Data martlar” veri ambarlarının alt kümeleridir. Veri ambarları bir iş probleminin tamamına yönelik bir bakış sağlarken, “data martlar” sadece belli bir iş probleminin özel bir alt kümesi için tasarlanırlar. İşletme üzerindeki karar vericilerin, işe ait tüm veriler üzerinde analiz yapmasına gerek olmayabilir. Bu kişiler sadece kendi birimleriyle ilgili verilere ulaşarak bunlara bağlı analizler yapmayı isteyebilirler, bu durumda veri ambarındaki tüm karmaşıklık içinde boğulmalarına gerek yoktur. Veri ambarlarının sadece bir konu kapsamında alt kümesini temsil eden data martlar, veri ambarları kadar ayrıntılı veri de barındırmazlar. Bu yüzden kolay anlaşılabilir ve yönlendirilebilirlerdir.

SQL (Yapısal Sorgu Dili): SQL, sorgu dili olarak bilinir. Analiz ve veri madenciliği ile karıştırmamak gerekmektedir. Sorgu ile sadece isteğe bağlı özel aranan yapılar veya sayıların ortaya çıkarılabileceği unutulmamalıdır.

OLAP (“Online Analytical Processing”): OLAP, veriler üzerinde çok boyutlu analizler yapılmasına olanak sağlayan bir yaklaşımdır. VM interaktif ve iteratif bir süreçtir. Anlamlı ilişkilerin ve örüntülerin keşfedilmesinde iş bilgisine sahip uzmanların deneyimlerinin ileri düzey yöntemlerle eklenmesi ile çıkarılabileceği unutulmaması gereken bir gerçektir. En önemlisi de, veri madenciliği bir sihir değildir.

2.1.3 Veri Madenciliği Uygulama Alanları

VM günümüz bilişim dünyasında çok farklı alanlarda başarılı bir şekilde uygulanmakta ve ciddi faydalar sağlamaktadır. Bunlardan bazıları,

- Hastalık ve kanser profili,
- Gen ve dizi analizi,
- Biyokimyasal sonuçların aykırılıklarının tespiti,
- Sahtekarlık içeren telefon veya kredi kartı kullanımlarının tespitine yönelik modellerin geliştirilmesi,
- Talep tahmini,
- Web sayfalarındaki bir sonraki sayfanın tahmini,
- Firmayı terk etme riski yüksek olan müşterinin tespiti,

- Müşterilerin kümelenmesi,
- Bir ürün veya hizmet alan müşterinin daha sonra hangi ürün veya hizmeti alacağını tespiti (2, 20).

Yukarıda bahsedilen başlıklara ilişkin bazı örnekler de aşağıda verilmiştir.

Anunciaçao, O., Gomes, B.C., Vinga, S., Gaspar, J., Oliverira, O.L. ve Rueff, J. 2010 yılında yaptıkları “A Data Mining Approach for the Detection of High-Risk Breast Cancer Groups” adlı çalışmada, meme kanseri için yüksek riskli grupların profillerini tespit etmişlerdir.

Chen, J., He, H., Jin, H., McAullay, D., Williams, G., Kelman, C. 2006 yılında yaptıkları “Identifying Risk Groups Associated with Colorectal Cancer” adlı çalışmada, kolorektal kanserinin tanımlanmasına yönelik çalışma yapmışlardır.

Huml, M., Silye, R., Zauner, G., Hutterer, S. ve Schilcher, K. 2013 yılında yaptıkları “Brain Tumor Classification Using AFM in Combination with Data Mining Techniques” adlı çalışmada, beyin kanseri tiplerinin sınıflandırılmasında veri madenciliği yöntemleri kullanmışlardır.

Xue, R., Li, J. ve Streveler, D.J. 2004 yılında yaptıkları “Microarray gene expression profile data mining model for clinical cancer research” adlı çalışmada, gen verileri üzerinde veri madenciliği uygulaması gerçekleştirmişlerdir.

Maes, S., Tuyls, K., Vanschoenwinkel, B. ve Manderick, B. 2002 yılında yaptıkları “Credit Card Fraud Detection Using Bayesian and Neural Networks” adlı çalışmada, kredi kartı sahtekarlığı için uygulama gerçekleştirmişlerdir.

Hilas, C. ve Mastorocostas, P. 2008 yılında yaptıkları “An application of supervised and unsupervised learning approaches to telecommunications fraud detection” adlı çalışmada, telekomünikasyon alanındaki sahtekarlık ile ilgili bir uygulama gerçekleştirmişlerdir.

Lu, X., Dong, Z.Y. ve Li, X. 2005 yılında yaptıkları “Electricity market price spike forecast with data mining techniques” adlı çalışmada, elektrik piyasası için enerji talep tahminlerini veri madenciliği yöntemleri ile modellemişlerdir.

Hung, S.Y., Yen, D.C. ve Wang, H.Y. 2006 yılında yaptıkları “Applying data mining to telecom churn management” adlı çalışmada, bir telekomünikasyon firmasına ait müşterilerin terk etme risklerini veri madenciliği yöntemleri ile analiz etmişlerdir.

Rygielski, C., Wang, J.C. ve Yen, D.C. 2002 yılında yaptıkları “Data mining techniques for customer relationship management” adlı çalışmada, analitik müşteri ilişkileri yönetimi, çapraz ve düz satış projeleri için veri madenciliği yöntemleri kullanmışlardır.

2.1.4 Veri Madenciliğinde Kullanılan Temel Metodoloji: CRISP-DM

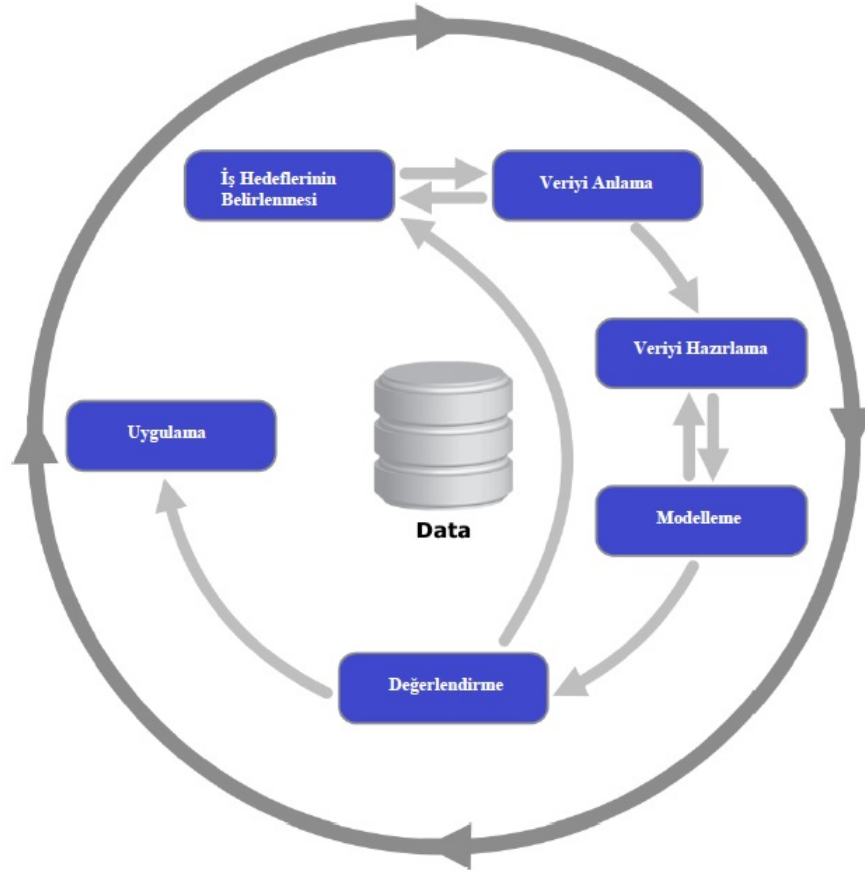
VM’nin birçok disiplini barındıran yapısı ve farklı uygulama alanlarındaki görevlerle prosedürlerin çeşitliliği, veri büyüklüğünden dolayı farklı ve kirli veri kaynakları ile çalışmadaki zorluklardan dolayı standart bir metodolojiye ihtiyaç duymaktadır.

VM projeleri öncesinde aşağıdaki soruların cevaplanması projelerin başarısı ve planlanması için yararlı olmaktadır.

- Nasıl bir iş problemi çözülmeye çalışılıyor?
- Hangi tür veri kaynakları mevcut ve bu iş problemi için ne çeşit verilere ihtiyaç duyulacak?
- Verileri analiz etmeden önce ne tür ön işleme ve veri temizleme işlemleri gerçekleştirilecek?
- Hangi veri madenciliği yöntemleri kullanılacak?
- Sonuçlar nasıl değerlendirilecek?

Bu tür sorular ve bir metodoloji ile VM projelerinde düzenli olarak ilerlenebilmektedir. “Cross Industry Standard Process for Data Mining” (CRISP-DM) metodolojisi, Daimler Chrysler AG, SPSS, NCR ve OHRA gibi lider veri madenciliği kullanıcıları ve tedarikçilerinden oluşan bir konsorsiyum tarafından geliştirilmiş ve dünyanın en büyük veri madenciliği çözümlerinin kullandığı altı adımdan oluşmuş bir süreçtir.

Bu altı adım aşağıdaki Şekil 2.3'te tanımlanmıştır.



Şekil 2.3. CRISP-DM Döngüsü

Bu altı adımı daha detaylı olarak tanımlamak gerekirse;

İş Hedeflerinin Belirlenmesi:

İlk adım proje amaçlarını ve iş gereksinimlerini anlamak, sonrasında da bu bilgiyi VM problemi tanımına ve amaçlara ulaşmak için oluşturulan ilk proje planına dönüştürme üzerine odaklanır. VM projesinin en zor kısmı olan bu adımda, ne tür bir analiz yapılması gerektiğinin kesinlikle iyi bir şekilde anlaşılması gerekir. Aksi takdirde tüm proje yanlışlıklar üzerine kurulmuş olacak ve bulunan sonuçlar amacı temsil etmeyecektir. Bu yüzden ilk ve en önemli adım, amacı açıkça belirlemek ve amaca giden süreci geliştirmektir. Amacı tanımlarken, neyi ölçmeye veya öngörmeye çalıştığınıza karar vermek gerekir.

Veriyi Anlama:

Veriyi anlama adımı öncelikle veriyi toplamakla başlar ve veri kümesinin

içinde hangi değişkenlerin olduğunun saptanması, bu değişkenlerin ve değerlerinin neleri ifade ettiklerini anlamakla devam eder. Eğer analist veriyi tanımıyor ise, projenin diğer aşamalarına geçmeden önce verilere hakim bir kişiden yardım alıp veriyi anlamalıdır. Aksi takdirde, yanlış bir model oluşturmak söz konusu olabilir.

Veriyi Hazırlama:

Veri hazırlığı aşaması, ham veriden veri madenciliği aracında kullanılacak en son veri kümesini (mining data) oluşturmak için yapılan tüm işlemleri kapsamaktadır. Veri madenciliğinin en önemli aşamalarından bir tanesi olan verinin hazırlanması aşaması analistin toplam zaman ve enerjisinin %70 - %80'ini harcamasına neden olmaktadır. Bu aşamadaki görevlerden bazıları; tablo oluşturma, kayıt ve değişken seçimi, veri temizliği, yeni değişkenler oluşturma ve modelleme araçları için verileri dönüştürme (transformation) işlemleridir.

Modelleme:

Bu adımda çeşitli modelleme teknikleri seçilip, uygulanır ve model parametreleri en uygun değerlere ayarlanır. Aynı tip VM problemleri için çeşitli teknikler mevcuttur. Bazı teknikler belli veri formatlarına ihtiyaç duymaktadır. Bu yüzden genellikle veri hazırlama adımına geri dönüş gerekir.

Klasik teknikler kullanıldığında tanımlanan problem için en uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu nedenle veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varıncaya kadar yinelenen bir süreçtir.

Model oluşturma süreci danışmanlı (supervised) ve danışmansız (unsupervised) öğrenmenin kullanıldığı modellere göre farklılık göstermektedir.

Öğrenme süreci tamamlandığında, tanımlanan kural cümleleri verilen yeni örneklerle uygulanır ve yeni örneklerin hangi sınıfa ait olduğu kurulan model tarafından belirlenir.

Değerlendirme:

Projenin bu aşamasında uygulamacının elinde kalitesi yüksek bir model mevcuttur. Modelin uygulama aşamasına geçmeden önce modelin eksiksiz olarak değerlendirilmesi ve iş amaçlarına uyup uymadığına emin olmak için model oluşturulana kadar yürütülen adımların tekrar gözden geçirilmesi büyük önem taşımaktadır. Buradaki temel amaç, analiz süresince gözden kaçan önemli bir noktanın var olup olmadığını belirlemektir.

Danışmanlı öğrenmede seçilen algoritmaya uygun olarak ilgili veriler hazırlandıktan sonra, ilk aşamada verinin bir kısmı modelin eğitimi, diğer kısmı

ise modelin geçerliliğinin test edilmesi için ayrılır. Modelin eğitimi, eğitim kümesi (training set) kullanılarak gerçekleştirildikten sonra, test kümesi (testing set) ile modelin doğruluk derecesi belirlenmektedir.

Uygulama:

En iyi modeli oluşturmak genellikle VM projesinin sonu değildir. Elde edilen bilginin düzenlenmesi ve müşterinin kullanacağı bir şekilde ifade edilmesi gerekmektedir. İhtiyaçlara göre uygulama adımı, bir rapor üretimi kadar basit veya oluşturulan modelin başka sistemlerin içine entegre edilmesi kadar karmaşık olabilir.

Birçok durumda, uygulama adımlarını gerçekleştirecek kişi veri analisti değil, kullanıcı olmaktadır. Kurulan modeller risk analizi, kredi değerlendirme, dolandırıcılık tespiti gibi işletme uygulamalarında doğrudan kullanılabilceği gibi, promosyon planlaması simülasyonuna entegre edilebilir veya tahmin edilen envanter düzeyleri yeniden sipariş noktasının altına düştüğünde, otomatik olarak sipariş verilmesini sağlayacak bir uygulamanın içine gömülebilir.

İzleme ve Güncelleme:

Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde değişiklikler ortaya çıkacaktır. Böyle bir durum karşısında modelin güncellenmesi gerekecektir. Günümüzde model güncelleme işleminin uzunca bir zaman alması ve iş yükü gerektirmesi nedeniyle, işletmelerin çoğu bu tarz bir çalışma yapmamakta ve oluşturulmuş modelleri uzun yıllar boyunca kullanmaktadır.

Sahtecilik tespit projesi için oluşturulan bir model düşünüldüğünde, bu modelin 1 yıl boyunca hatta 2 ay boyunca bile sürekli kullanılmaması gerektiği bir gerçektir. Bunun sebebi, sahtekarların yakalandıkça taktiklerini değiştirmeye başlayacak olmalarıdır. Dolayısıyla yeni taktiklerle gelen yeni sahtekarların profilleri, oluşturulan modelde bulunan sahtekar profiline uymayacak ve model bu sahtekarları yakalayamaz duruma gelecektir. Böyle bir durum ile karşılaşmamak için sahtekarlık modellerinin belirli aralıklarla güncellenmesi gerekmektedir.

Sonuç olarak güncelliğini yitirmiş modellerin tespiti için bu modellerin sürekli olarak izlenmesi gerekmektedir. Tahmin edilen ve gözlenen değişkenler arasındaki farklılığı gösteren grafikler model sonuçlarının izlenmesinde kullanılan yararlı bir yöntemdir (20).

2.1.5 Veri Madenciliğinde Kullanılan Modeller

VM literatüründe modelleme yöntemleri temel olarak,

- Tanımlayıcı (Descriptive) ve
- Çıkarımsal (Inferential)

olmak üzere iki gruba ayrılmaktadırlar.

Tanımlayıcı modellerde amaç, veri kümesindeki değişkenler arasındaki ilişkileri özetleyen örüntüleri ortaya çıkartmakken, tahmine dayalı modellerde belirli bir değişkenin değerini diğer değişkenlerden yararlanarak tahmin etmek amaçlanır. Regresyon analizi kullanılarak elde edilen sonuçları göz önüne aldığımızda; istatistiksel olarak anlamlı değişkenler bize hangi değişken veya değişkenlerin bağımlı değişken üzerinde etki ettiğini göstermektedir. Bu bilgi tanımlayıcı olarak değerlendirilebilir. Kurulan modelin, gelecek için yaptığı kestirimler ise tahmin edici olarak değerlendirilebilir.

Birliktelik kuralları, kümeleme analizi, faktör analizi, anormal kayıtların tespiti gibi yöntemler tanımlayıcı modelleme yöntemleri arasında yer almaktadır. Regresyon analizi, karar ağaçları, yapay sinir ağları, DVM, bayesci ağlar, k-en yakın komşular gibi yöntemler ise çıkarımsal modelleme yöntemleri arasında yer almaktadır (21).

Öğrenme Yöntemleri

VM’de modelleme algoritmaları öğrenme yöntemlerine bağlı olarak,

- Danışmanlı (Supervised) ve
- Danışmansız (Unsupervised)

öğrenme olmak üzere ikiye ayrılmaktadırlar.

Danışmanlı öğrenme yöntemlerinde, bağımlı değişken bir dizi bağımsız değişken tarafından tahmin edilmektedir. Regresyon analizi, karar ağaçları gibi yöntemler bu gruba örnek olarak verilebilir. Danışmansız öğrenme yöntemlerinde ise, bağımlı değişken bulunmamaktadır. Faktör analizi, kümeleme analizi gibi yöntemler bu gruba örnek olarak verilebilir.

Kullanım Alanı

VM’de modelleme algoritmaları kullanım alanına bağlı olarak,

- Sınıflama ve Regresyon,
- Kümeleme,
- Birliktelik Kuralları

olmak üzere üçe ayrılmaktadırlar.

Sınıflama ve regresyon yöntemleri, tahmin edici; kümeleme ve birliktelik kuralları yöntemleri ise tanımlayıcı amaçlı kullanılan modelleme yöntemleri içerisinde yer almaktadır (4, 21).

Sınıflama ve Regresyon:

İstenilen bir değişken bağımlı değişken ve diğerleri tahmin edici (bağımsız) değişkenler olarak adlandırılır. Amaç, girdi olarak tahmin edici değişkenlerin yer aldığı modelde, çıktının bağımlı değişkenin değerinin bulunduğu anlamlı bir model kurmaktır. Bağımlı değişken sayısal değil ise problem sınıflama problemidir. Eğer bağımlı değişken sayısal ise problem regresyon problemi olarak adlandırılır.

Mevcut verilerden hareket ederek geleceğin tahmin edilmesinde faydalanan ve VM yöntemleri içerisinde en yaygın kullanıma sahip olan sınıflama ve regresyon modelleri arasındaki temel fark, tahmin edilen bağımlı değişkenin kategorik veya sürekli bir değere sahip olmasıdır (2).

Kümeleme:

Kümeleme işlemi, heterojen yapıya sahip bir kitleyi daha homojen birkaç alt gruba ya da kümeye bölme işlemidir. Sınıflama ile kümelemeyi birbirinden ayıran en önemli fark, kümeleme işleminin sınıflama işleminde olduğu gibi önceden belirlenmiş bir takım sınıflara göre bölme yapmamasıdır. Sınıflamada her bir veri, önceden sınıflandırılmış bir takım sınıflar üzerinde yapılan bir eğitim neticesinde ortaya çıkan bir modele göre önceden belirlenmiş olan bir sınıfa atanmaktadır. Kümeleme işleminde ise önceden tanımlanmış sınıflar ya da örnek sınıflar bulunmamaktadır. Verilerin kümelenmesi işlemi, verilerin birbirlerine olan benzerliklerine göre yapılmaktadır. Oluşan sınıfların hangi anlamları taşıdığının belirlenmesi tamamen çözümlemeyi yapan kişiye kalmıştır.

Kümeleme işlemi çoğunlukla bir başka VM uygulaması için bir ilk işlem olarak kullanılır. Kümeleme modellerinde amaç, küme üyelerinin birbirlerine çok benzediği, ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve

veri tabanındaki kayıtların bu farklı kümelere bölünmesidir (2).

Birliktelik Kuralları:

Birliktelik kuralları, VM arařtırmalarında çok büyük yatırımlar yapılan, VM'nin özel bir uygulama alanıdır. Birliktelik kuralları aynı işlem içinde çoğunlukla beraber görülen nesneleri içeren kurallardır. Birliktelik kurallarının bulunması ile pazar sepeti çözümlemesi yapılmaktadır. Pazar sepeti çözümlemesinde, nesneler müşteriler tarafından satın alınan ürünlerdir ve bir işlem (kayıt) ise birçok nesneyi içinde bulunduran tek bir satın almadır. Pazar sepeti çözümlemesinde sıklıkla beraber alınan nesneler üzerine çalışılır. Bulunan kurallar ile nesnelerin birbiri ile nasıl ilişkili olduđu bilgisine ulaşılır. Bir alışveriş sırasında veya birbirini izleyen alışverişlerde müşterinin hangi mal veya hizmetleri satın alma eğiliminde olduğunun belirlenmesi, müşteriye daha fazla ürün satma yollarından birisidir (2).

Tez kapsamında sınıflama yöntemleri arasındaki başarılar karşılaştırılacağından, bu konu detaylı olarak bir sonraki bölümde ele alınmıştır.

2.2 Sınıflama Yöntemleri

Sınıflama ve regresyon terimleri, danışmanlı öğrenme yöntemleri arasında yer alan ve ileriye yönelik çıkarımlarda kullanılan yöntemleri ifade etmektedir. Sınıflama, kategorik bir bağımlı değişkenin tahmini için kullanılırken; regresyon terimi veri madenciliği literatüründe, sürekli bir bağımlı değişkenin tahmini için kullanılmaktadır. Sınıflama, bağımlı değişkenin kategorik olduđu tahmine dayalı yöntemlerdir. Bu yöntemlerde bağımlı değişken “sınıf değişkeni” olarak da adlandırılabilir. Sınıflama yöntemlerinde amaç, sınıflar arasındaki sınırların belirlenerek bağımsız değişkenin verilen değerlerine göre bağımlı değişken değerini doğru bir şekilde tahmin etmektir. Literatürde en çok kullanılan sınıflama yöntemlerine örnek olarak, karar ağaçları, lojistik regresyon, bayesci ağlar, sinir ağları, k-en yakın komşular, destek vektör makinaları verilebilir (4, 21).

Bu tez kapsamında kullanılan yöntemler sadece sınıflama olduğundan, regresyon yöntemleri bu kapsamda ele alınmamıştır. Tez kapsamında, bağımlı değişken iki durumlu olarak ele alınmış ve aşağıdaki sınıflama algoritmaları kullanılmıştır.

1. Karar Ağaçları

- Chi-squared Automatic Interaction Detection (CHAID)
- Classification and Regression Tree (C&RT)
- C5.0

2. Bayesci Ağlar

- Tree Augmented Naive Bayes (TAN)
- Markov Blanket (MB)

3. Destek Vektör Makinaları (DVM)

Aşağıdaki bölümlerde, tez kapsamında kullanılan sınıflama yöntemleri ile ilgili bilgi verilmiştir.

2.2.1 Karar Ağaçları

Karar Ağaçları, adından da anlaşılabilirceği gibi ağaç görünümünde tahmin edici bir yöntemdir. Karar ağacı yöntemini kullanarak verinin sınıflanması iki basamaklı bir işlemdir. İlk basamak öğrenme basamağıdır. Öğrenme basamağında önceden bilinen bir eğitim verisi, model oluşturmak amacı ile sınıflama algoritması tarafından çözümlenir. Öğrenilen model, sınıflama kuralları veya karar ağacı olarak gösterilir. İkinci basamak ise sınıflama basamağıdır. Sınıflama basamağında test verisi, sınıflama kurallarının veya karar ağacının doğruluğunu belirlemek amacıyla kullanılır. Eğer doğruluk kabul edilebilir oranda ise, kurallar yeni verilerin sınıflandırılmasında kullanılır. Karar ağaçlarının anlaşılması ve değiştirilmesi kolaydır. Karar ağaçları ile oluşturulan model, bir ‘karar kuralları’ kümesi içerir. Bu kurallar kümesi sınıf değişkeninin tahmin edilmesinde kullanılır.

Karar ağaçları, veri madenciliğinde en çok tercih edilen modellerden birisidir, çünkü diğer yöntemlere oranla daha hızlı bir şekilde model sonucu vermektir. Bu modellerin diğer bir üstünlüğü, basit bir yapıda olmaları ve anlaşılabilirliklerinin kolay olmasıdır (4, 8, 20).

Birbirinden genel olarak bağımsız çalışmalarla ortaya çıkmış olan karar ağaçları teorisi için bazı gelişim adımları aşağıda verilmiştir;

- 1962 yılında Hunt, yapısal bilgiler kazanmak amacıyla kullanmıştır,
- 1963 yılında Feingbaum ve Simon, diskriminant ağları olarak kullanmıştır,

- 1967 yılında Samuel, oyunlarda kendini geliştirebilmek amacıyla kullanmıştır,
- 1969 yılında Quinlan, problem çözümlemelerinde kullanmıştır,
- 1980 yılında Kass, CHAID (ki-kare otomatik etkileşim tespiti) algoritmasını geliştirmiştir,
- 1983 yılında Carbonell, Michalski ve Mitchell, en basit algoritma mantığı olan TDIDT (Top-Down Induction of Decision Trees algoritmasını geliştirmişlerdir,
- 1984 yılında Breiman, Friedman, Stone ve Olshen, C&RT (sınıflama ve regresyon ağacı) algoritmasını geliştirmişlerdir,
- 1986 yılında Quinlan, ID3 (iterative dichotomiser 3) algoritmasını, TDIDT algoritmasına dayalı bir şekilde geliştirmiştir,
- 1987 yılında Riverst R. Boolean, Decision List algoritmasını geliştirmiştir,
- 1993 yılında Quinlan, ID3 algoritmasını geliştirerek C4.5 algoritmasını elde etmiştir,
- 1998 yılında Quinlan, C4.5 algoritmasını geliştirerek C5.0 algoritmasını elde etmiştir.

Karar ağaçları, bağımlı değişken iki kategorili olsa da teori olarak n tane sınıfa örnekleme dağıtacak ölçütler yaratan bir yöntemdir. İlk olarak, her sınıfın gözlemleri için en iyi kırılmayı sağlayan bağımsız değişkeni bilgi kuramı, olasılık teorisi, istatistiksel yöntemler, optimizasyon ve yapay zeka algoritmaları yardımıyla seçer. O değişken üzerinden koşullandırma yani dal oluşturarak koşulu sağlayan gözlemleri çocuk düğüme taşınır. Bu şekilde kök düğümden çocuk düğümler elde edilir. Düğüm oluşturma işlemi algoritmanın özelliğine göre belirlenmiş durumlar sağlanana (Örneğin, CHAID için istatistiksel olarak önemsiz bulunana kadar) kadar devam eder.

Ağacın tipine ya da yapısına göre belirlenen durdurma kuralı devreye girene kadar ya da ağaçta daha fazla kırılma yapılamayacak duruma gelene kadar ebeveyn düğümler, çocuk düğümlere ayrılır. Kırılma bittiğinde sadece çocuk düğüm olan ve terminal düğüm olarak adlandırılan düğümler yorumlanır. Bu düğümler kurallara göre artık kırılmayacak haldedir. Kırılma için çeşitli ölçütler veya durdurma seçenekleri bulunmaktadır.

Ağacın Büyümesi

Ağaç oluşturulurken ilk amaç, elde bulunan örneklemdaki tüm gözlemleri n sınıfa bölmektir. Her sınıfın bireylerini en iyi şekilde kırarak olan değişkenin nasıl seçileceği bilinmelidir. En önemli bağımsız değişkenin bulunup gözlemlerin bu değişken üzerinden k düğüme ayrılması temel amaçtır. Uygun kırılma sayısı bağımsız değişkenin tipine bağlı olarak elde edilir. İkili kategorili bir değişken tek koşul oluşturur ve kırılma ölçütünün sağlanıp sağlanmadığına göre değişir. n kategorili değişkenler $(n-1)$ koşul oluşturarak n ya da daha az kırılma oluşturur. n 'den az kırılma, kategorilerin birleştirilerek aynı düğüme düşmesiyle ortaya çıkar. Sürekli değişkenler ise $x_1, x_2, \dots, x_n$ şeklinde sıralanır ve $x_1 \leq x_2 \leq \dots \leq x_n$ için kırılma ölçütü $x < ort(x_1, x_{n+1})$ şeklinde tanımlanır.

Ağacın oluşumunda değişkenlerin tiplerinin önemli olduğu kırılma kurallarından görülmektedir. Dolayısıyla değişken tiplerini paket program kullanırken sürekli ya da kategorik olarak belirlemenin önemli olduğu unutulmamalıdır. En uygun kırılma bulunduğu kırılma gerçekleşir ve işlem ayrışmaları her düğüm için durdurma ölçütü aktif hale gelene kadar devam ettirilir.

Büyümenin Durdurulması

Düğümlerdeki gözlemlerin özdeş ve ayrılmayacak duruma gelmesi, her düğüme sadece bir sınıfa ait olan gözlemlerin bulunması gibi durumlar düğümün ve dolayısıyla ağacın daha fazla ilerlemesini engeller. Ancak bu durum genellikle teorik sonuç olup, ağacın derinliğinin daha fazla artmasını engellemek için kullanılan durdurma ölçütleridir. Durdurma ölçütü genellikle aşağıdaki durumları kapsamaktadır;

- Ağacın derinliği belirlenen limite ulaştığında,
- Her düğüme bulunan gözlem sayısı belirlenenden daha az sayıya düştüğünde,
- Ağaç kırılmada istenilen başarıya ulaştığında.

Budama

Ağaçlar, derinlik arttıkça düğümlerde çok az birey bulundurmaya başlar. Bağımlı değişken bu yüzden bütün örüntüler ile ilişkili gibi gözükabilir. İstatistiksel güvenilirlik için düğümdaki gözlem sayıları yorumlanabilecek kadar olmalıdır. Sınıflandırma hataları da budama yapılmasını gerektirir. Budama aynı

zamanda aşırı uyum ve aşırı öğrenme (overfitting, overtraining) oluşumunu da engelleyecektir.

Algoritma maksimum boyutta ağacı oluşturup budama yapılırken optimal budama eşik değerini belirler. Bu belirleme iki şekilde yapılmaktadır;

1. Örneklem büyüklüğü yeterli ise örneklemden bir test örnekleme yaratılır. Maksimum ağacın, her alt ağacında yaratılmış olan test örnekleme kullanılır ve en iyi budanmış ağaç en az hata oranına sahip olan alt ağaç olur.
2. Örneklem yeterli değilse çapraz geçerlilik gerekir. Bütün olası alt ağaçlar için hata oranları bulunur ve birleştirilerek en iyi ağaç seçilir, budama yapılır.

Tez kapsamında kullanılan CHAID, C&RT ve C5.0 yöntemlerine ilişkin bazı bilgiler aşağıdaki bölümlerde verilmiştir. Karar ağaçları ile ilgili detaylı bilgi, Han, J., Kamber, M. (4) ve Larose, D.T. (17)'de bulunabilir.

CHAID

1980 yılında G.V. Kass tarafından geliştirilen CHAID yöntemi hem sınıflama hem de regresyon amacı ile kullanılmaktadır. Bağımlı değişkeni en fazla etkileyen bağımsız değişken seçiminde; bağımlı değişkenin kategorik olması durumunda Ki-Kare testi kullanılırken, sürekli olması durumunda F testi kullanılır. Kayıp değerleri ayrıca bir kategoride değerlendirmektedir. Kategorik ve sürekli değişkenler üzerinde çalışabilmesi ve ağaçta her dalın ikiden fazla alt gruba ayırabilmesi gibi özelliklerinden dolayı günümüzde de tercih edilen bir algoritmadır (4, 17).

C&RT

1984 yılında Breiman, Friedman, Olshen ve Stone tarafından geliştirilen C&RT yöntemi de CHAID gibi hem sınıflama hem de regresyon amacı ile kullanılmaktadır. C&RT yönteminde her dal ikili alt gruplara ayrılarak büyümektedir. Ayırım işlemi kategorik bağımlı değişkenler için gini, twoing ve sıralı twoing, sürekli değişkenler için ise en küçük kare sapması (Least-Squared Deviation) kullanılmaktadır. Bu hesaplamalarda kar (profit), yanlış sınıflama maliyeti (misclassification costs) ve değişken kategorileri arasındaki önceliklerin tanımlanabilmesi (priors) gibi çeşitli esneklikler sağlamaktadır. Kayıp değerler için vekil bir değişken belirlemekte ve kayıp değerleri doldurmaktadır (4, 17).

C5.0

C4.5 ve C5.0 yöntemleri Quinlan tarafından, ID3 yönteminin eksik yönlerini gidermek için geliştirilmiştir. C5.0 yöntemi, 1992 yılında Quinlan tarafından C4.5 yönteminin geliştirilmiş olarak önerilmiştir. Diğer yöntemlerden farklı olarak bilgi kazanımı (information gain) ölçütü ile ağacı büyütmektedir. C5.0 yönteminin C4.5 yönteminden; çalışma hızı, etkin bellek kullanımı, yanlış sınıflama maliyeti, boosting, çapraz geçerlilik (cross-validation), değişken eleme seçenekleri gibi avantajları vardır. C5.0 yöntemi sadece sınıflama amacı ile kullanılabilir. Kayıp değerleri ağaç büyütme hesaplamalarında dikkate almamaktadır (4, 17).

2.2.2 Bayesci Ağlar

Bayes teoremini temel alan ve 1985 yılında Judea Pearl tarafından geliştirilen Bayesci Ağlar, inanç ağları (belief network) olarak da bilinmektedir. Bayesci ağlar, olasılıksal grafik modelleri ailesinin bir üyesidir. Olasılıksal grafik modelleri, ilgilenilen problemin kesin olmayan tanım kümesi hakkındaki bilgiyi temsil etmek için kullanılır. Bu grafiklerde düğümler (nodes) rastlantı değişkenlerini, düğümler arasındaki bağlar (edges) ise rastlantı değişkenleri arasındaki olasılıksal bağımlılık durumlarını gösterir. Bu bağımlılıklar genellikle bilinen istatistiksel ve sayısal yöntemlerden yararlanılarak analiz edilir. Değişkenler arasındaki bağların yönlendirilmemiş olduğu grafiksel modeller genellikle “Markov rastgele alanları” ya da “Markov ağları” olarak adlandırılır. Bu ağlar, “Markov örtüsü” kavramından yararlanarak iki düğüm arasındaki bağımsızlığın basit bir tanımını yapar. Değişkenler arasındaki bağların yönlendirilmiş olduğu grafiksel modeller ise bayesci ağlar olarak adlandırılır. Bayesci ağlardaki değişkenler arasında koşullu bağımsızlık (conditional independence) özelliği geçerlidir. Bu, naive bayes öğrenmenin temel varsayımlarından birisini oluşturmaktadır. Koşullu bağımsızlık, bir değişkenin ebeveynlerinin durumları bilindiğinde, bu değişkenin, soy değişkenleri dışındaki değişkenlerden bağımsız olduğu anlamına gelmektedir. Bu özellik, değişkenlerin çok değişkenli olasılık dağılımının hesaplanmasında kullanılan parametre sayısının azalmasını sağlar (22, 23).

Bayesci ağ modelinin bağımlılık yapısı ve koşullu olasılık fonksiyonları veriden elde edilir. Bu süreç öğrenme olarak adlandırılır. Amaç, mevcut veriler kullanılarak, doğruluğu yüksek en iyi modeli elde etmektir. Öğrenmenin iki farklı çeşidi vardır. Grafikte yer alan bağlantıları belirleyen bağımlılık yapısını öğrenme yani yapısal öğrenme ve koşullu olasılık fonksiyonlarının parametrik yapısını

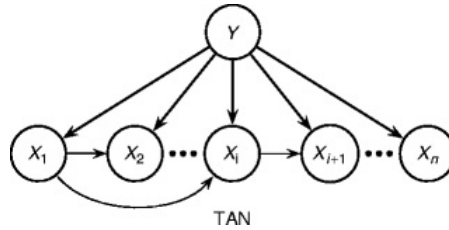
öğrenme yani parametre öğrenmedir.

Bayeci ağların en önemli özelliği neden sonuç ilişkisini değil de modellediği dünyayı doğrudan temsil etmesidir. Ağ üzerinde çizilen oklar gerçek ilişkisel nedenler olup, çıkarsama yapılırken izlenen yollar değildirler. Bu yönüyle de kural tabanlı veya yapay sinir ağları ile oluşturulan modellerden farklıdırlar (16, 23).

Tez kapsamında kullanılan TAN ve MB yöntemlerine ilişkin bazı bilgiler aşağıdaki bölümlerde verilmiştir. Bayesci ağlar ile ilgili detaylı bilgi, Ersel, D. (22), Ben-Gal, I. (23), Cheng, J., Bell, D., Liu, W. (24) ve Larose, D.T. (16)'de bulunabilir.

TAN

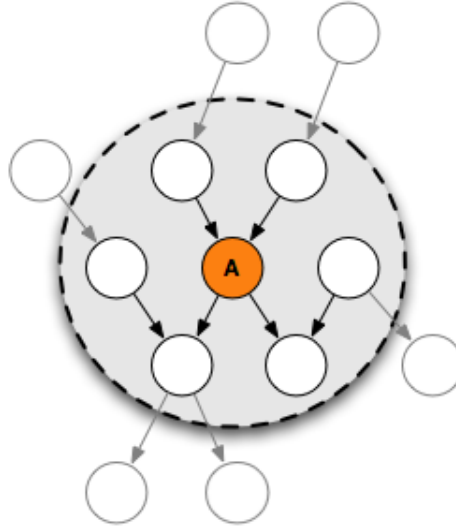
TAN, yarı naive bayesci öğrenme yöntemleri içerisinde yer alan ve basit bir bayesci ağ oluşturan bir yöntemdir. Naive bayesci öğrenme yöntemlerinde koşullu bağımsızlık gibi katı bir varsayım var iken yarı naive bayes öğrenme yöntemlerinde bu varsayım esnetilebilmektedir. Her bir bağımsız değişkenin bir diğer bağımsız değişkene ve bağımlı değişkene bağlı olduğu bir yapıya izin vermektedir. Bu yapı sınıflamanın doğruluğunu arttırmaktadır (22, 23). Aşağıdaki Şekil 2.4'te TAN örneği verilmiştir.



Şekil 2.4. TAN Örneği

MB

MB yöntemi, ebeveynlerin, çocukların ve çocukların ebeveynlerinin diğer bütün değişkenler ile koşullu bağımsız olduğunu varsaymaktadır. TAN'dan farklı olarak değişken seçiminde kullanılan bazı özellikler bu yöntemlere adapte edilmiştir (16, 23). Aşağıdaki Şekil 2.5'te MB örneği verilmiştir.



Şekil 2.5. MB Örneği

2.2.3 Destek Vektör Makinaları

DVM'ler kavram olarak ilk kez 1960'larda ortaya atılmıştır. Yaklaşık 50 yıla yakın bir geçmişe sahip olan bu yöntem, son yıllarda bir çok alanda sağladığı başarılı sınıflama performansından dolayı tercih edilmektedir. Genelleme hatasını en aza indirmeye ve veriyi ayıracak doğru ya da düzlemin marjin uzaklığını en yüksek yapma mantığına dayanarak sınıflama yapmaktadır.

DVM'lerin temelini oluşturan düşüncelerin yer aldığı ilk kitap 1982 yılında Vapnik tarafından hazırlanmıştır. DVM'nin sahip olduğu matematiksel algoritmalar başlangıçta iki sınıflı doğrusal verilerin sınıflandırılması problemi için tasarlanmış, daha sonra çok sınıflı ve doğrusal olmayan verilerin sınıflandırılması ve hatta sürekli bağımlı değişkenler için genelleştirilmiştir. DVM'nin çalışma prensibi iki sınıflı birbirinden ayırabilen en uygun karar fonksiyonun tahmin edilmesi, başka bir ifadeyle iki sınıflı birbirinden en uygun şekilde ayırabilen çoklu düzlemin tanımlanması esasına dayanmaktadır. Veriler doğrusal olarak ayrılabilen ve doğrusal olarak ayrılamayan olmak üzere iki grupta ele alınmaktadır. Her iki durumda da çoklu düzlemler belirlenmektedir. Doğrusal olarak ayrılamayan veriler özellik uzayı olarak tanımlanan yüksek boyutlu bir uzayda görüntülenir. Böylece verilerin doğrusal olarak ayrımı yapılabilen ve sınıflar arasındaki çoklu düzlem belirlenebilmektedir. Çekirdek fonksiyonları yardımı ile verinin daha yüksek bir boyuta dönüştürülmesi sağlanmaktadır. Tez kapsamında RBF adı verilen çekirdek fonksiyon kullanılmıştır (4, 25, 26).

RBF

Yapay sinir ağı kökenli radial basis function (RBF), çekirdek fonksiyonlar arasında kullanımı en yaygın çekirdek fonksiyondur. Aşağıdaki eşitlik ile hesaplanmaktadır.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma > 0 \quad (2.1)$$

Eşitlikteki karesel ifade, iki vektör arasındaki karesel öklid uzaklığı olarak düşünülebilir. Gama değeri ise yarıçap kontrolünü sağlayan parametredir.

2.3 Model Başarı Ölçütleri

Sınıflama algoritmalarının veriyi ne kadar doğru sınıfladığını ölçmede kullanılan çeşitli model başarı ölçütleri bulunmaktadır. Bu tez kapsamında literatürde sıklıkla kullanılan;

- Verinin İki Alt Örnekleme Ayrılması (Hold-Out Yöntemi)
- Genel Doğruluk (Overall Accuracy)
- Dengeli Doğruluk (Balanced Accuracy)
- Duyarlılık (Sensitivity)
- Seçicilik (Specificity)
- Matthews Korelasyon Katsayısı (Matthews Correlation Coefficient)
- F-Ölçütü (F-Measure)

ölçütleri kullanılmıştır.

Yukarıda söz edilen model başarı ölçütleri, iki durumlu bağımlı değişkenin (örneğin, 0 ve 1 gibi) olduğu durum için aşağıda verilen 2×2 'lik sınıflama tablosunun gözeleri kullanılarak hesaplanmaktadır.

Tablo 2.1. Sınıflama Tablosu

		Gerçek Durum	
		1	0
Model sonucunda tahmin edilen	1	Doğru Pozitif (DP)	Yanlış Pozitif (YP)
	0	Yanlış Negatif (YN)	Doğru Negatif (DN)

2.3.1 Verinin İki Alt Örneklemeye Ayrılması - Hold-Out Yöntemi

Hold-Out yöntemi, analiz edilecek verinin rastgele olarak en az iki alt örneklemeye ayrılması mantığını temel almaktadır. Algoritmanın parametreleri belirlediği yani eğitildiği bölüme, eğitim kümesi (training set), sonuçlarının test edildiği kümeye ise, test kümesi (testing set) adı verilmektedir. Algoritmanın eğitimi, eğitim kümesi kullanılarak tamamlandıktan sonra, test kümesi ile modelin doğruluk değeri (accuracy) belirlenmektedir.

Bir modelin doğruluğunun test edilmesi için kullanılan en basit yöntem, basit geçerlilik (Simple Validation). Bu yöntemde genelde verilerin %5'i ile %33'ü arasındaki bir kısmı test kümesi olarak ayrılır ve kalan kısım üzerinde algoritmanın eğitimi gerçekleştirildikten sonra, bu veriler üzerinde test işlemi yapılır. Bir sınıflama modelinde, yanlış olarak sınıflanan gözlemlerin sayısı, toplam gözlem sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan gözlem sayısının toplam gözlem sayısına bölünmesi ile ise doğruluk oranı hesaplanmaktadır.

Eğitim ve test kümesi için genellikle kullanılan oranlar, (0.90 - 0.10), (0.85 - 0.15), (0.80 - 0.20), (0.75 - 0.25) ve (0.70 - 0.30)'dur. Bu oranlar, analiz edilecek verinin büyüklüğüne ve yapısına göre değişim göstermektedir. Tez kapsamında bu oranlar, eğitim kümesi %80 ve test kümesi %20 olacak şekilde düzenlenmiştir. Yorumlar ve çıktılarda, test kümesinin sonuçları kullanılmıştır.

2.3.2 Genel Doğruluk

Literatürde sıklıkla kullanılan genel doğruluk (GD - overall accuracy) ölçütünün, bağımlı değişken prevelansının dengesiz olduğu durumlarda doğru yorumlanacak bir sonuç vermediği çeşitli makale ve kitaplarda belirtilmiş ve bu gibi durumlarda farklı başarı ölçütlerinin kullanılması gerektiği vurgulanmıştır (4, 27–29).

Model başarı ölçütleri bölümünde incelenen birçok başarı ölçütünün, bağımlı değişken prevelansının dengesiz olduğu durumlarda model başarısını doğru bir şekilde göstermemesinden dolayı, model karşılaştırmalarında literatürde önerilen dengeli doğruluk (DD) ve Mathew korelasyon katsayısı (MKK) ölçütleri kullanılmıştır. Literatürde verilen bilgilere paralel olarak, niçin bazı ölçütlerin kullanılmaması gerektiğine yönelik sonuçlardan bazıları sadece tartışmaya katkı

sağlaması açısından Bölüm 4’da ele alınacaktır. Aşağıdaki eşitlikte yer alan N , örneklem büyüklüğünü ifade etmektedir.

Aşağıdaki eşitlik ile hesaplanmaktadır;

$$GD = \frac{DP + DN}{N} \quad (2.2)$$

2.3.3 Dengeli Doğruluk

Bağımlı değişken kategorilerinin dengesiz olduğu durumlarda kullanılması önerilen bir diğer başarı ölçütü ise DD ölçütüdür (30, 31).

Aşağıdaki eşitlik ile hesaplanmaktadır;

$$DD = \frac{\frac{DP}{(DP+YN)} + \frac{DN}{(YP+DN)}}{2} \quad (2.3)$$

2.3.4 Duyarlılık

Duyarlılık (DUY) en basit anlamda gerçek pozitifler arasında modelin pozitifleri doğru bir şekilde sınıflama başarısıdır ve aşağıdaki eşitlik ile elde edilir (32).

$$DUY = \frac{DP}{(DP + YN)} \quad (2.4)$$

2.3.5 Seçicilik

Seçicilik (SEC) ise gerçek negatifler arasında modelin negatifleri doğru bir şekilde sınıflama başarısıdır ve aşağıdaki eşitlik ile hesaplanmaktadır (32).

$$SEC = \frac{DN}{(YP + DN)} \quad (2.5)$$

2.3.6 Matthew Korelasyon Katsayısı

Veri madenciliği çalışmalarında iki durumlu bağımlı değişken yapısı altında model başarısını ölçmede kullanılan bir ölçüttür. Avantajlı yanı, bağımlı değişkene ait prevelans değerinin dengesiz olduğu durumlarda diğer başarı ölçütlerine göre daha doğru sonuç vermesidir. -1 ile +1 arasında değer almaktadır (33-35).

+1 değeri; en iyi durumu, 0 değeri; rastgele bir tahmin yapıldığını, -1 değeri ise ters bir tahmin yapıldığını göstermektedir. İstatistik literatüründe phi katsayısı olarak da bilinmektedir.

Aşağıdaki eşitlik ile hesaplanmaktadır;

$$MKK = \frac{(DP \times DN) + (YP \times YN)}{\sqrt{(DP + YP)(DP + YN)(DN + YP)(DN + YN)}} \quad (2.6)$$

2.3.7 F-Ölçütü

DUY ve SEC ölçütlerinin tek başına yeterli olmadıkları düşünülmüş ve elde edilmiştir. DUY ve pozitif kestirim değeri ($DP / (DP + YP)$) ölçütlerinin (Recall and Precision) harmonik ortalamasıdır. F ölçütünün doğru negatif hücrelerini dikkate almadığı göz önüne alınmalıdır. Bu durumdan dolayı, Matthews ilişki katsayısı gibi ölçütler önerilmiştir (33).

Aşağıdaki eşitlik ile hesaplanmaktadır;

$$F = \frac{2DP}{(2DP + YP + YN)} \quad (2.7)$$

3 GEREÇ VE YÖNTEM

3.1 Benzetim Çalışması

Çalışmada, bağımlı ve bağımsız değişkenlerin, örneklem büyüklüğünün, değişken sayısının ve bağımlı değişken prevelansının farklı yapılarda olduğu durumlarda, Karar Ağaçları, Bayesci Ağlar ve DVM sınıflama yöntemlerinin model başarılarının karşılaştırılması için Monte Carlo benzetiminden yararlanılmıştır. Veri türetme aşamasında Tablo 3.1’de verilen $3^5 = 243$ farklı senaryo için veriler üretilmiştir.

Tablo 3.1’de verilen benzetim senaryolarının genel çerçevesi ise aşağıdaki gibi oluşturulmuştur:

- Örneklem büyüklüğünün, sınıflama yöntemlerinin model başarılarının üzerine olan etkisini değerlendirmek için 3 farklı örneklem büyüklüğü senaryosu ($N = 250$, $N = 1000$ ve $N = 5000$) kullanılmıştır.
- Bağımsız değişkenler arasındaki ilişki yapısının model başarıları üzerine olan etkisini değerlendirmek için 3 farklı ilişki yapısı kullanılmıştır. Buna göre bağımsız değişkenler arasındaki düşük ilişki ($r = 0.10 - 0.30$), orta düzeyde ilişki ($r = 0.31 - 0.59$) ve yüksek ilişki ($r = 0.60 - 1$) yapıları, bağımsız değişkenlerin normal dağılım gösterdiği varsayımı altında; ortalaması 0, varyansı da 1 olan çok değişkenli normal dağılımdan üretilmiştir. Belirtilen ilişki yapılarına sahip çok değişkenli normal dağılımlı veri üretmek için Box-Muller yaklaşımı (36) kullanılmış ve ilişki matrisi girdi olarak tanımlanmıştır.
- Bağımsız değişkenler ile bağımlı değişken arasındaki ilişki yapısının model başarılarının üzerine olan etkisini değerlendirmek için 3 farklı ilişki yapısı kullanılmıştır. Buna göre bağımsız değişkenler ile bağımlı değişken arasında düşük ilişki ($r = 0.10 - 0.30$), orta düzeyde ilişki ($r = 0.31 - 0.59$) ve yüksek ilişki ($r = 0.60 - 1$) yapıları ele alınmıştır.

Örneğin; bağımsız değişkenler arasında yüksek ilişki ve bağımsız değişkenler ile bağımlı değişken arasında düşük ilişki olan senaryo için çok değişkenli normal dağılımda aşağıdaki kitle ilişki matrisi tanımlanmıştır:

$$\begin{bmatrix} & X_1 & X_2 & X_3 & X_4 & X_5 & Y \\ X_1 & 1 & 0.80 & 0.80 & 0.80 & 0.80 & 0.20 \\ X_2 & 0.80 & 1 & 0.80 & 0.80 & 0.80 & 0.20 \\ X_3 & 0.80 & 0.80 & 1 & 0.80 & 0.80 & 0.20 \\ X_4 & 0.80 & 0.80 & 0.80 & 1 & 0.80 & 0.20 \\ X_5 & 0.80 & 0.80 & 0.80 & 0.80 & 1 & 0.20 \\ Y & 0.20 & 0.20 & 0.20 & 0.20 & 0.20 & 1 \end{bmatrix}$$

Yukarıdaki ilişki matrisinde bağımsız değişkenler $(X_1, X_2, \dots, X_5)$ arasındaki (kitle) ilişki değerinin 0.80 olarak tanımlanmasının nedeni, ilişki yapısının yüksek olduğu sınırlar olan $[0.61-1]$ aralığı için ilişki değerlerinin uniform dağılım gösterdiği varsayımı altında beklenen değeri olmasıdır. Benzer şekilde $X_1, X_2, \dots, X_5$ bağımsız değişkenleri ile bağımlı değişken Y arasındaki ilişki değerinin 0.20 olarak belirlenmesinin sebebi de düşük ilişki aralığının $[0.10-0.30]$ beklenen değeri olmasıdır.

Benzetim çalışmasının temelini oluşturan rastgeleliği sağlamak ve her bir senaryo tekrarında ilgili aralıktan rastgele ilişki katsayıları elde etmek için ise $[a-b]$ aralığında uniform dağılımlı ilişki değerleri üretilmiştir. Bu sayede çok değişkenli normal dağılımda kitle ilişki matrisi yukarıdaki gibi iken, 1000 tekrardan elde edilecek örneklem ilişki matrislerinin farklı olması sağlanmıştır. Örneğin; herhangi bir tekrar için elde edilen örneklem ilişki matrisi aşağıdaki gibidir:

$$\begin{bmatrix} & X_1 & X_2 & X_3 & X_4 & X_5 & Y \\ X_1 & 1 & 0.693 & 0.799 & 0.680 & 0.912 & 0.107 \\ X_2 & 0.693 & 1 & 0.897 & 0.796 & 0.697 & 0.193 \\ X_3 & 0.799 & 0.897 & 1 & 0.797 & 0.777 & 0.294 \\ X_4 & 0.680 & 0.796 & 0.797 & 1 & 0.798 & 0.196 \\ X_5 & 0.912 & 0.697 & 0.777 & 0.798 & 1 & 0.292 \\ Y & 0.107 & 0.193 & 0.294 & 0.196 & 0.292 & 1 \end{bmatrix}$$

Benzetim çalışmasında ilişkili veri yapısı üretilirken sadece pozitif ilişki yapısı dikkate alınmıştır. Negatif ilişki değerleri dikkate alınmamıştır.

- Tez çalışmasında sınıflama yöntemlerinin iki kategoride sınıflama yaptığı durumlar, diğer bir ifade ile bağımlı değişkenin iki kategorili olduğu durumlar ele alınmıştır. Bağımlı değişkenin dağılım yapısının model başarılarına olan etkilerini görebilmek için ise iki durumlu bağımlı değişkenin prevalansı 3 farklı şekilde ele alınmıştır. Düşük prevalans için $P(Y = 1) = 0.10$, orta düzeyde prevalans için $P(Y = 1) = 0.50$ ve yüksek prevalans için $P(Y = 1) = 0.75$ olarak ele alınmıştır.

Verilen ilişki yapısı ile çok değişkenli normal dağılımdan Y bağımlı değişkeni sürekli sayısal bir veri olarak üretilecektir. Sürekli sayısal yapıda ve standart normal dağılımı gösteren bağımlı değişken, standart normal dağılımı için ilgili z tablo değerleri kullanılarak ilgili prevalansı sağlayacak şekilde iki kategoriye aşağıdaki gibi ayrılmıştır:

Düşük prevalansa sahip iki durumlu bağımlı değişken (Y');

$$Y' = \begin{cases} 1 & \text{Eğer } Y \geq 1.29 \\ 0 & \text{Eğer } Y < 1.29. \end{cases}$$

olarak oluşturulmuştur. Benzer şekilde orta düzeyde prevalansa sahip iki durumlu bağımlı değişken üretilemek için

$$Y' = \begin{cases} 1 & \text{Eğer } Y \geq 0 \\ 0 & \text{Eğer } Y < 0. \end{cases}$$

ve yüksek prevalans için ise,

$$Y' = \begin{cases} 1 & \text{Eğer } Y \geq -0.675 \\ 0 & \text{Eğer } Y < -0.675. \end{cases}$$

şeklinde oluşturulmuştur. Kullanılan 1.29, 0 ve -0.675 değerleri standart normal dağılımda sırasıyla 0.10, 0.50 ve 0.75 olasılıkları için kritik değerlerdir.

- Bağımsız değişken sayısının modellerin sınıflama performansına etkisini değerlendirmek amacıyla bağımsız değişken sayısı $k = 5$, $k = 10$ ve $k = 25$ olarak alınmıştır.

Her bir senaryo için 1000 tekrar yapılmış ve üretilen veriler karar ağacı yöntemlerinden CHAID, C&RT, C5.0, bayesci ağlar yöntemlerinden TAN, MB ve DVM yöntemlerinden RBF ile analiz edilerek bu yöntemlerin sınıflama

başarılarına ilişkin sonuçlar elde edilmiştir. Yöntemlere ilişkin detaylı bilgi Bölüm 2.2’de ve hangi senaryonun hangi bileşenlerden oluştuğu bilgisi ise Bölüm 4’de verilmiştir.

Model başarılarını ölçmek için ise detayları Bölüm 2.3’te verilen ve literatürde sıklıkla kullanılan başarı ölçütlerinden; Hold-out Yöntemi, F, GD, DD, DUY, SEC, MKK ölçütleri ve grafiksel yöntemler kullanılmıştır.

Belirlenen senaryolar için benzetim çalışması IBM SPSS Statistics 19 paket program ile yazılan syntax ve scriptler kullanılarak yapılmıştır. Elde edilen veriler ise IBM SPSS Modeler 14.1 paket program yardımıyla analiz edilmiştir. Benzetim çalışması ve modellemeye ait kodlar Ek 1 ve Ek 2’de verilmiştir.

Tablo 3.1. Tez çalışmasında kullanılan benzetim senaryoları

Ana Senaryo	Alt Senaryo	Senaryo Düzeyleri
Örneklem Büyüklüğü (N)	Küçük örneklem	$N = 250$
	Orta genişlikte örneklem	$N = 1000$
	Büyük örneklem	$N = 5000$
Bağımsız Değişkenler Arasındaki İlişki Yapısı	Düşük ilişki	$r = 0.10 - 0.30$
	Orta düzeyde ilişki	$r = 0.31 - 0.59$
	Yüksek ilişki	$r = 0.60 - 1$
Bağımsız Değişkenler ile Bağımlı Değişken Arasındaki İlişki Yapısı	Düşük ilişki	$r = 0.10 - 0.30$
	Orta düzeyde ilişki	$r = 0.31 - 0.59$
	Yüksek ilişki	$r = 0.60 - 1$
Bağımlı Değişkenin Dağılım Yapısı (Prevelans)	Düşük prevelans	$P(Y = 1) = 0.10$
	Orta düzeyde prevelans	$P(Y = 1) = 0.50$
	Yüksek prevelans	$P(Y = 1) = 0.75$
Modeldeki Bağımsız Değişken Sayısı (k)	Az	$k = 5$
	Orta	$k = 10$
	Çok	$k = 25$

4 BULGULAR

Karşılaştırılan modellere ilişkin sonuçlar aşağıda Bölüm 4.1 ve Bölüm 4.2’de yer almaktadır. Bölüm 4.1, senaryo bazlı, tez kapsamında verilen bütün başarı kriterlerinin ortalamalarını içermektedir. Bölüm 4.2 ise sadece MKK ve DD başarı kriterlerine ilişkin grafiksel sonuçları içermektedir.

Bölüm 4.1’de detaylı olarak niçin bütün başarı kriterlerinin yorumlamalarda kullanılmadığı anlatılmıştır.

4.1 Tablo Bazlı Sonuçlar

Aşağıdaki Tablo 4.1’de her bir senaryo ve algoritmaya ilişkin, DD başarı kriterine göre 1000 tekrarın ortalama değerleri yer almaktadır. SNR kolonu; senaryo numarasını, N kolonu; gözlem sayısını, BSIZ_K kolonu; bağımsız değişkenler arasındaki ilişki miktarını (D: Düşük, O: Orta, Y: Yüksek), PRE kolonu; prevelans değerini (D: Düşük, O: Orta, Y: Yüksek), DS kolonu; bağımsız değişken sayısını ve BLI_K kolonu; bağımsız değişkenler ile bağımlı değişken arasındaki ilişki miktarını (D: Düşük, O: Orta, Y: Yüksek) göstermektedir.

Örneğin 1. satır, 1 nolu senaryoya ilişkin DD başarı kriterine göre C5.0, CHAID, C&RT, TAN, MB, DVM algoritmasına ait sonuçları ifade etmektedir. 1 nolu senaryonun 1000 tekrarın DD ortalaması; C5.0 için 0.4556, CHAID için 0.2786, C&RT için 0.4649, TAN için 0.5166, MB için 0.4608 ve DVM için 0.4743 olarak hesaplanmıştır.

Tablo 4.1, Tablo 4.2, Tablo 4.3, Tablo 4.4, Tablo 4.5 ve Tablo 4.6’da yer alan 6 nolu senaryo dikkatle incelendiğinde, Bölüm 4.2’de niçin sadece MKK ve DD kriterlerin kullanıldığının cevabı bulunacaktır. Tablo 4.4’te yer alan GD değerinin 0.901 gibi yüksek bir değer hesaplandığı yani model başarısının iyi olduğu düşünülmektedir. Ancak, Tablo 4.5’te yer alan DUY değerinin 0 olarak hesaplandığı görülmektedir. Duyarlılığın 0 çıkması, bağımlı değişken kategorilerinin dengesizliğinden kaynaklanan ciddi bir sıkıntıyı işaret etmekte ve modelin başarısının sanıldığı gibi iyi olmadığını göstermektedir. Bu durum, bağımlı değişken kategorilerinin dengesiz olduğu yapılarda, GD, SEC, DUY ve hatta F ölçütlerinin kullanımının niçin önerilmediğini göstermektedir (30, 31).

Aşağıda verilmiş olan Tablo 4.1, Tablo 4.2, Tablo 4.3, Tablo 4.4, Tablo 4.5 ve Tablo 4.6 kullanıcılara hangi senaryoda hangi yöntemin kullanılması gerektiğiyle ilgili model başarı ölçütleri ışığında öneriler de vermektedir. Örneğin, 6 nolu senaryo için önerilen yöntemin hem MKK hem de DD kriterlerine göre TAN olduğu söylenebilmektedir. Ayrıca Tablo 4.1, Tablo 4.2, Tablo 4.3, Tablo 4.4, Tablo 4.5 ve Tablo 4.6'da yer alan senaryolarda başarı kriteri en yüksek olan değer koyu renk ile ifade edilmiştir.

Tablo 4.1. Yöntemlerin Senaryo Bazlı DD Kriterine Göre Sonuçları

SNR	N	BSIZ_K	PRE	DS	BLI_K	C5.0	CHAID	C&RT	TAN	MB	DVM
1	250	D	D	5	D	0,4556	0,2786	0,4649	0,5166	0,4608	0,4743
2	1000	D	D	5	D	0,4503	0,4450	0,4610	0,5462	0,4529	0,4553
3	5000	D	D	5	D	0,4525	0,4510	0,4522	0,5842	0,5386	0,4513
4	250	O	D	5	D	0,4587	0,2732	0,4366	0,5142	0,4609	0,4599
5	1000	O	D	5	D	0,4540	0,4457	0,4290	0,5305	0,4538	0,4553
6	5000	O	D	5	D	0,4507	0,4507	0,4485	0,5581	0,4589	0,4507
7	250	Y	D	5	D	0,4575	0,2330	0,4188	0,4934	0,4577	0,4525
8	1000	Y	D	5	D	0,4530	0,4239	0,4156	0,5136	0,4514	0,4529
9	5000	Y	D	5	D	0,4507	0,4507	0,4462	0,5207	0,4507	0,4515
10	250	D	O	5	D	0,5241	0,5057	0,5489	0,5561	0,4259	0,5959
11	1000	D	O	5	D	0,5716	0,5733	0,5726	0,5770	0,5168	0,6032
12	5000	D	O	5	D	0,5829	0,5823	0,5851	0,5975	0,6006	0,6072
13	250	O	O	5	D	0,4952	0,4481	0,5290	0,5406	0,3862	0,5684
14	1000	O	O	5	D	0,5671	0,5667	0,5587	0,5606	0,3155	0,5816
15	5000	O	O	5	D	0,5700	0,5716	0,5713	0,5779	0,5580	0,5847
16	250	Y	O	5	D	0,4617	0,3780	0,5258	0,5248	0,3614	0,5587
17	1000	Y	O	5	D	0,5549	0,5635	0,5474	0,5482	0,2546	0,567
18	5000	Y	O	5	D	0,5651	0,5645	0,5624	0,5618	0,2762	0,5686
19	250	D	Y	5	D	0,4456	0,3682	0,5377	0,5424	0,4418	0,5429
20	1000	D	Y	5	D	0,4417	0,4671	0,5595	0,5804	0,4561	0,5537
21	5000	D	Y	5	D	0,4396	0,5301	0,5611	0,6119	0,6092	0,6175
22	250	O	Y	5	D	0,4230	0,3333	0,5211	0,5179	0,4323	0,5109
23	1000	O	Y	5	D	0,4019	0,4159	0,5317	0,5600	0,3956	0,4900
24	5000	O	Y	5	D	0,3881	0,4224	0,4990	0,5873	0,5162	0,5136
25	250	Y	Y	5	D	0,4170	0,3070	0,5218	0,5085	0,4208	0,4759
26	1000	Y	Y	5	D	0,3868	0,3878	0,4993	0,5492	0,3786	0,4442
27	5000	Y	Y	5	D	0,3774	0,3774	0,4544	0,5751	0,3821	0,4283
28	250	D	D	10	D	0,4919	0,4029	0,5144	0,5282	0,4687	0,5334
29	1000	D	D	10	D	0,4877	0,4688	0,4705	0,5600	0,4574	0,5323
30	5000	D	D	10	D	0,4694	0,4538	0,4572	0,5892	0,4898	0,4900
31	250	O	D	10	D	0,4815	0,3371	0,4754	0,5159	0,4661	0,5202
32	1000	O	D	10	D	0,4656	0,4551	0,4438	0,5383	0,4528	0,5148
33	5000	O	D	10	D	0,4520	0,4510	0,4484	0,5614	0,4513	0,4529
34	250	Y	D	10	D	0,4721	0,2592	0,4503	0,5067	0,4613	0,4890
35	1000	Y	D	10	D	0,4620	0,4400	0,4065	0,5172	0,4514	0,4544
36	5000	Y	D	10	D	0,4507	0,4507	0,4479	0,5404	0,4507	0,4507
37	250	D	O	10	D	0,5578	0,5498	0,5568	0,5678	0,4755	0,5994
38	1000	D	O	10	D	0,5816	0,5764	0,5793	0,5917	0,5431	0,6169
39	5000	D	O	10	D	0,5876	0,5838	0,5875	0,6113	0,5877	0,6206
40	250	O	O	10	D	0,5267	0,5135	0,5347	0,5392	0,3916	0,5651
41	1000	O	O	10	D	0,5668	0,5652	0,5565	0,5638	0,3169	0,5804
42	5000	O	O	10	D	0,5721	0,5714	0,5723	0,5820	0,4987	0,5884
43	250	Y	O	10	D	0,4903	0,4265	0,5211	0,5207	0,3884	0,5585
44	1000	Y	O	10	D	0,5648	0,5630	0,5456	0,5430	0,2505	0,5647
45	5000	Y	O	10	D	0,5659	0,5637	0,5614	0,5595	0,2548	0,5701
46	250	D	Y	10	D	0,5070	0,4751	0,5601	0,5554	0,4807	0,5747
47	1000	D	Y	10	D	0,5220	0,5300	0,5718	0,5823	0,4725	0,6095
48	5000	D	Y	10	D	0,5547	0,5821	0,5889	0,6114	0,5849	0,6520
49	250	O	Y	10	D	0,4663	0,4042	0,5379	0,5380	0,4375	0,5405
50	1000	O	Y	10	D	0,4428	0,4522	0,5425	0,5580	0,3927	0,5769
51	5000	O	Y	10	D	0,4203	0,4701	0,5265	0,5794	0,4475	0,6070
52	250	Y	Y	10	D	0,4400	0,3313	0,5258	0,5152	0,4295	0,5383
53	1000	Y	Y	10	D	0,4097	0,4025	0,5079	0,5399	0,3754	0,5401
54	5000	Y	Y	10	D	0,3822	0,3813	0,4669	0,5595	0,3769	0,5182
55	250	D	D	25	D	0,5242	0,4911	0,5321	0,5394	0,4853	0,5361
56	1000	D	D	25	D	0,5335	0,5190	0,4831	0,5559	0,4610	0,5435
57	5000	D	D	25	D	0,5186	0,4657	0,4698	0,5769	0,4954	0,5612

Tablo 4.1 – *Devami*

SNR	N	BSIZ_K	PRE	DS	BLIK	C5.0	CHAID	C&RT	TAN	MB	DVM
58	250	O	D	25	D	0,5090	0,4511	0,4964	0,5257	0,4673	0,5132
59	1000	O	D	25	D	0,4979	0,4748	0,4225	0,5371	0,4517	0,5199
60	5000	O	D	25	D	0,4600	0,4508	0,4508	0,5456	0,4517	0,5321
61	250	Y	D	25	D	0,4847	0,3415	0,4583	0,5019	0,4531	0,5138
62	1000	Y	D	25	D	0,4707	0,4532	0,4160	0,5210	0,4504	0,5156
63	5000	Y	D	25	D	0,4533	0,4492	0,4450	0,5301	0,4508	0,5365
64	250	D	O	25	D	0,5642	0,5663	0,5587	0,5851	0,4952	0,5944
65	1000	D	O	25	D	0,5754	0,5774	0,5796	0,6076	0,5522	0,6229
66	5000	D	O	25	D	0,5879	0,5852	0,5882	0,6269	0,4864	0,6312
67	250	O	O	25	D	0,5474	0,5403	0,5276	0,5509	0,3967	0,5546
68	1000	O	O	25	D	0,5697	0,5647	0,5584	0,5730	0,2885	0,5776
69	5000	O	O	25	D	0,5738	0,5702	0,5728	0,5873	0,3876	0,5908
70	250	Y	O	25	D	0,5079	0,4957	0,5204	0,5257	0,2708	0,5473
71	1000	Y	O	25	D	0,5628	0,5613	0,5418	0,5454	0,2510	0,5605
72	5000	Y	O	25	D	0,5652	0,5602	0,5596	0,5581	0,2484	0,5690
73	250	D	Y	25	D	0,5384	0,5362	0,5503	0,5735	0,4938	0,5698
74	1000	D	Y	25	D	0,5677	0,5623	0,5786	0,5930	0,5093	0,6082
75	5000	D	Y	25	D	0,5922	0,5932	0,5966	0,6133	0,5322	0,6509
76	250	O	Y	25	D	0,5053	0,4764	0,5324	0,5415	0,4387	0,5446
77	1000	O	Y	25	D	0,5158	0,5127	0,553	0,5613	0,3863	0,5615
78	5000	O	Y	25	D	0,4895	0,5280	0,5427	0,5742	0,3742	0,5900
79	250	Y	Y	25	D	0,4634	0,3983	0,5232	0,5213	0,3842	0,5270
80	1000	Y	Y	25	D	0,4496	0,4484	0,5124	0,5369	0,3764	0,5437
81	5000	Y	Y	25	D	0,4017	0,406	0,4668	0,5497	0,3751	0,5627
82	250	D	D	5	O	0,6179	0,5579	0,6383	0,6461	0,5765	0,7000
83	1000	D	D	5	O	0,6794	0,6575	0,6852	0,7157	0,6600	0,7568
84	5000	D	D	5	O	0,7257	0,7113	0,7135	0,7441	0,7298	0,7849
85	250	O	D	5	O	0,5461	0,4907	0,6022	0,5994	0,5464	0,5729
86	1000	O	D	5	O	0,5655	0,5276	0,6220	0,6514	0,5637	0,5729
87	5000	O	D	5	O	0,6197	0,5966	0,6362	0,6775	0,6439	0,7388
88	250	Y	D	5	O	0,4984	0,4507	0,5651	0,5690	0,5405	0,4816
89	1000	Y	D	5	O	0,4883	0,4586	0,5638	0,6150	0,4731	0,4855
90	5000	Y	D	5	O	0,4717	0,4511	0,5286	0,6472	0,5954	0,4884
91	250	D	O	5	O	0,6908	0,6806	0,6903	0,7119	0,6002	0,7567
92	1000	D	O	5	O	0,7154	0,7027	0,7142	0,7425	0,6999	0,7665
93	5000	D	O	5	O	0,7314	0,7144	0,7241	0,7507	0,7534	0,7696
94	250	O	O	5	O	0,6632	0,6622	0,6439	0,6645	0,6523	0,6957
95	1000	O	O	5	O	0,6722	0,6691	0,6696	0,6853	0,6818	0,7009
96	5000	O	O	5	O	0,6837	0,6791	0,6843	0,6942	0,6991	0,7051
97	250	Y	O	5	O	0,6522	0,6492	0,6084	0,6180	0,6463	0,6548
98	1000	Y	O	5	O	0,6511	0,6497	0,6399	0,6457	0,5766	0,6603
99	5000	Y	O	5	O	0,6539	0,6545	0,6547	0,6528	0,6604	0,6636
100	250	D	Y	5	O	0,6871	0,6665	0,6847	0,6977	0,575	0,7565
101	1000	D	Y	5	O	0,7115	0,7004	0,7102	0,7401	0,6499	0,7718
102	5000	D	Y	5	O	0,7339	0,7159	0,7267	0,7528	0,7505	0,7771
103	250	O	Y	5	O	0,6286	0,6137	0,6296	0,6490	0,6236	0,6918
104	1000	O	Y	5	O	0,6788	0,6726	0,671	0,6805	0,6580	0,7184
105	5000	O	Y	5	O	0,6957	0,6883	0,6945	0,6924	0,684	0,7264
106	250	Y	Y	5	O	0,5382	0,5544	0,5949	0,6097	0,6118	0,639
107	1000	Y	Y	5	O	0,6184	0,6313	0,6453	0,6457	0,5558	0,6857
108	5000	Y	Y	5	O	0,6752	0,6762	0,6725	0,6543	0,6431	0,6987
109	250	D	D	10	O	0,6608	0,6185	0,6559	0,6751	0,5303	0,7608
110	1000	D	D	10	O	0,6897	0,6816	0,6875	0,7384	0,5228	0,8057
111	5000	D	D	10	O	0,7235	0,7072	0,7125	0,7566	0,6877	0,8222
112	250	O	D	10	O	0,5851	0,5259	0,6102	0,6065	0,5193	0,622
113	1000	O	D	10	O	0,6236	0,5984	0,6457	0,6345	0,517	0,6973
114	5000	O	D	10	O	0,6798	0,6672	0,6601	0,6483	0,6308	0,7564
115	250	Y	D	10	O	0,5250	0,4754	0,5692	0,5634	0,5281	0,5709
116	1000	Y	D	10	O	0,5240	0,4639	0,5701	0,5927	0,4673	0,5894
117	5000	Y	D	10	O	0,5061	0,4524	0,5518	0,6035	0,5748	0,6156
118	250	D	O	10	O	0,7127	0,6909	0,7109	0,7602	0,2999	0,8039
119	1000	D	O	10	O	0,7308	0,7097	0,7221	0,7921	0,5709	0,8174
120	5000	D	O	10	O	0,7500	0,7172	0,7288	0,8037	0,8075	0,8224
121	250	O	O	10	O	0,6610	0,6598	0,6458	0,6743	0,5751	0,6991
122	1000	O	O	10	O	0,6723	0,6678	0,6731	0,6979	0,588	0,7105
123	5000	O	O	10	O	0,6868	0,6801	0,6864	0,7080	0,7142	0,7171
124	250	Y	O	10	O	0,6466	0,647	0,5998	0,6150	0,6427	0,6551
125	1000	Y	O	10	O	0,6506	0,646	0,6344	0,6418	0,5321	0,6596
126	5000	Y	O	10	O	0,6544	0,6536	0,6538	0,6474	0,663	0,6644
127	250	D	Y	10	O	0,6977	0,6717	0,694	0,7363	0,2746	0,7911
128	1000	D	Y	10	O	0,7203	0,7004	0,7154	0,7774	0,5511	0,8165
129	5000	D	Y	10	O	0,7425	0,7149	0,7270	0,7924	0,7867	0,8250
130	250	O	Y	10	O	0,6432	0,6355	0,6390	0,6539	0,5611	0,6866
131	1000	O	Y	10	O	0,6723	0,6681	0,6714	0,6805	0,5806	0,7208

Tablo 4.1 – Devamı

SNR	N	BSIZ_K	PRE	DS	B_LI_K	C5.0	CHAID	C&RT	TAN	MB	DVM
132	5000	O	Y	10	O	0,6928	0,6862	0,6926	0,6908	0,6875	0,7369
133	250	Y	Y	10	O	0,5757	0,5632	0,5922	0,6011	0,6038	0,6356
134	1000	Y	Y	10	O	0,6350	0,6443	0,6394	0,6288	0,5044	0,6857
135	5000	Y	Y	10	O	0,6771	0,6752	0,6718	0,6371	0,6397	0,7014
136	250	D	D	25	O	0,6717	0,6477	0,6592	0,7275	0,4900	0,7885
137	1000	D	D	25	O	0,6962	0,6855	0,6870	0,7877	0,4811	0,839
138	5000	D	D	25	O	0,7218	0,7010	0,7059	0,7738	0,4554	0,8692
139	250	O	D	25	O	0,6059	0,5799	0,6178	0,617	0,5134	0,6092
140	1000	O	D	25	O	0,6309	0,6388	0,6463	0,6323	0,5134	0,6666
141	5000	O	D	25	O	0,6720	0,6918	0,6743	0,6285	0,6213	0,7442
142	250	Y	D	25	O	0,5512	0,5004	0,5695	0,5621	0,5070	0,5680
143	1000	Y	D	25	O	0,5622	0,4910	0,5792	0,5801	0,4604	0,6111
144	5000	Y	D	25	O	0,5550	0,4586	0,5547	0,5850	0,5404	0,6770
145	250	D	O	25	O	0,7249	0,7021	0,7169	0,8213	0,2758	0,8440
146	1000	D	O	25	O	0,7451	0,7197	0,7254	0,8490	0,2498	0,8711
147	5000	D	O	25	O	0,7633	0,7173	0,7302	0,8652	0,8692	0,8806
148	250	O	O	25	O	0,6485	0,6514	0,6486	0,6981	0,3614	0,6910
149	1000	O	O	25	O	0,6663	0,6683	0,6775	0,7151	0,3150	0,7148
150	5000	O	O	25	O	0,6835	0,6800	0,6845	0,7188	0,7243	0,7234
151	250	Y	O	25	O	0,6386	0,6414	0,5946	0,618	0,635	0,6473
152	1000	Y	O	25	O	0,6468	0,6425	0,6314	0,6452	0,5052	0,6604
153	5000	Y	O	25	O	0,6524	0,6506	0,6517	0,6439	0,6651	0,6635
154	250	D	Y	25	O	0,7070	0,6823	0,7039	0,8055	0,1984	0,8302
155	1000	D	Y	25	O	0,7310	0,7082	0,7164	0,8283	0,1251	0,8652
156	5000	D	Y	25	O	0,7521	0,7144	0,7265	0,8395	0,7172	0,8785
157	250	O	Y	25	O	0,6376	0,6319	0,6372	0,6722	0,3481	0,6622
158	1000	O	Y	25	O	0,6586	0,6631	0,6732	0,6884	0,3623	0,7138
159	5000	O	Y	25	O	0,6851	0,6845	0,6914	0,6891	0,6901	0,7435
160	250	Y	Y	25	O	0,6080	0,5817	0,5948	0,6008	0,5859	0,6324
161	1000	Y	Y	25	O	0,6450	0,6407	0,6333	0,6225	0,4869	0,6749
162	5000	Y	Y	25	O	0,6754	0,6688	0,6683	0,6233	0,6372	0,6986
169	250	Y	D	5	Y	0,7577	0,7317	0,7528	0,752	0,7304	0,8031
170	1000	Y	D	5	Y	0,8026	0,8134	0,8013	0,7744	0,7274	0,8328
171	5000	Y	D	5	Y	0,8188	0,8220	0,8159	0,7633	0,7254	0,8404
178	250	Y	O	5	Y	0,8011	0,8012	0,7904	0,8095	0,8308	0,8358
179	1000	Y	O	5	Y	0,8143	0,8123	0,8143	0,8183	0,8283	0,8357
180	5000	Y	O	5	Y	0,8230	0,8179	0,8223	0,8182	0,8260	0,8374
187	250	Y	Y	5	Y	0,7969	0,7972	0,7825	0,7912	0,8002	0,8272
188	1000	Y	Y	5	Y	0,8142	0,8114	0,8105	0,8098	0,8014	0,8372
189	5000	Y	Y	5	Y	0,8235	0,8178	0,8227	0,8125	0,8051	0,8403
196	250	Y	D	10	Y	0,7565	0,7659	0,7599	0,7323	0,7194	0,7897
197	1000	Y	D	10	Y	0,7915	0,8059	0,7954	0,7425	0,7155	0,8313
198	5000	Y	D	10	Y	0,8154	0,8193	0,8131	0,7371	0,7155	0,8459
205	250	Y	O	10	Y	0,8032	0,8029	0,794	0,8122	0,8418	0,8369
206	1000	Y	O	10	Y	0,8158	0,8146	0,8168	0,8252	0,8409	0,8431
207	5000	Y	O	10	Y	0,8245	0,8194	0,8235	0,8217	0,8371	0,8436
214	250	Y	Y	10	Y	0,7913	0,7944	0,7862	0,7896	0,8037	0,8270
215	1000	Y	Y	10	Y	0,8089	0,807	0,8109	0,8052	0,8029	0,8393
216	5000	Y	Y	10	Y	0,8226	0,8168	0,8223	0,8052	0,8061	0,8458
223	250	Y	D	25	Y	0,7544	0,7770	0,7703	0,7314	0,7163	0,7794
224	1000	Y	D	25	Y	0,7726	0,8000	0,7945	0,7319	0,7084	0,8191
225	5000	Y	D	25	Y	0,8064	0,8135	0,8098	0,7173	0,7071	0,8448
232	250	Y	O	25	Y	0,7942	0,7993	0,7982	0,812	0,8456	0,8322
233	1000	Y	O	25	Y	0,8066	0,8137	0,8172	0,8321	0,8477	0,8430
234	5000	Y	O	25	Y	0,8222	0,8188	0,8238	0,8288	0,8454	0,8463
241	250	Y	Y	25	Y	0,7835	0,7888	0,7905	0,785	0,805	0,8197
242	1000	Y	Y	25	Y	0,7950	0,8047	0,8118	0,8043	0,8046	0,8390
243	5000	Y	Y	25	Y	0,8192	0,8163	0,8224	0,8002	0,8070	0,8502

Tablo 4.2. Yöntemlerin Senaryo Bazlı MKK Kriterine Göre Sonuçları

SNR	N	BSIZ_K	PRE	DS	B_LI_K	C5.0	CHAID	C&RT	TAN	MB	DVM
1	250	D	D	5	D	0,0065	0,0048	0,0683	0,0699	0,0115	0,0207
2	1000	D	D	5	D	0,0057	0,0079	0,0285	0,0596	0,0014	0,0026
3	5000	D	D	5	D	0,0006	0,0001	0,001	0,0447	0,0389	0,0001
4	250	O	D	5	D	0,0085	0,0044	0,0516	0,0678	0,0087	0,0077
5	1000	O	D	5	D	0,0022	0,0014	0,0168	0,0487	0,0019	0,0018
6	5000	O	D	5	D	0	0	0,0002	0,0333	0,0041	0
7	250	Y	D	5	D	0,0062	0,0042	0,0412	0,0392	0,0081	0,0024
8	1000	Y	D	5	D	0,0013	0,0004	0,0111	0,0353	0,0002	0,0008
9	5000	Y	D	5	D	0	0	0	0,0197	0	0,0002
10	250	D	O	5	D	0,1117	0,1204	0,1112	0,1243	0,0642	0,1943
11	1000	D	O	5	D	0,1359	0,1407	0,1423	0,1536	0,102	0,2058

Tablo 4.2 – Devami

SNR	N	BSIZ_K	PRE	DS	BLI_K	C5.0	CHAIID	C&RT	TAN	MB	DVM
12	5000	D	O	5	D	0,1634	0,163	0,1682	0,1948	0,2009	0,2141
13	250	O	O	5	D	0,0963	0,0944	0,0853	0,1015	0,0364	0,1496
14	1000	O	O	5	D	0,124	0,1264	0,1146	0,1216	0,0118	0,1623
15	5000	O	O	5	D	0,1365	0,1411	0,1401	0,1555	0,1424	0,1688
16	250	Y	O	5	D	0,0835	0,0821	0,0802	0,0802	0,0291	0,1332
17	1000	Y	O	5	D	0,1095	0,119	0,0927	0,0986	0,0016	0,1328
18	5000	Y	O	5	D	0,1229	0,1253	0,121	0,1221	0,0065	0,1361
19	250	D	Y	5	D	0,0429	0,0379	0,09	0,103	0,0363	0,0954
20	1000	D	Y	5	D	0,0298	0,0414	0,0867	0,1048	0,0289	0,0681
21	5000	D	Y	5	D	0,0225	0,056	0,0733	0,1144	0,1429	0,0693
22	250	O	Y	5	D	0,0297	0,0335	0,0752	0,0759	0,0283	0,0724
23	1000	O	Y	5	D	0,0105	0,0162	0,0594	0,0786	0,0054	0,0359
24	5000	O	Y	5	D	0,0039	0,0147	0,0387	0,0779	0,0737	0,0252
25	250	Y	Y	5	D	0,0283	0,0294	0,0746	0,0634	0,0197	0,05
26	1000	Y	Y	5	D	0,0045	0,0051	0,0417	0,0604	0,0008	0,019
27	5000	Y	Y	5	D	0,0006	0,0006	0,0213	0,048	0,0012	0,0077
28	250	D	D	10	D	0,045	0,0261	0,1012	0,0898	0,018	0,0978
29	1000	D	D	10	D	0,025	0,0132	0,0444	0,0917	0,0038	0,0445
30	5000	D	D	10	D	0,007	0,0013	0,0049	0,1011	0,0229	0,0087
31	250	O	D	10	D	0,0335	0,0098	0,0643	0,0773	0,0151	0,0793
32	1000	O	D	10	D	0,01	0,004	0,0255	0,0658	0,0007	0,0336
33	5000	O	D	10	D	0,0004	0,0001	0,0005	0,0659	0,0002	0,0005
34	250	Y	D	10	D	0,0213	0,0047	0,0511	0,0628	0,0097	0,0372
35	1000	Y	D	10	D	0,0075	0,0008	0,0148	0,0427	0,0004	0,0017
36	5000	Y	D	10	D	0	0	0,0001	0,0357	0	0
37	250	D	O	10	D	0,1284	0,1328	0,126	0,1479	0,0827	0,201
38	1000	D	O	10	D	0,1569	0,1488	0,1558	0,1832	0,1176	0,2331
39	5000	D	O	10	D	0,1735	0,1666	0,1733	0,2226	0,1738	0,241
40	250	O	O	10	D	0,1076	0,1166	0,0949	0,1026	0,04	0,1402
41	1000	O	O	10	D	0,1245	0,1249	0,1114	0,1283	0,0103	0,1598
42	5000	O	O	10	D	0,1412	0,141	0,1422	0,1638	0,1	0,176
43	250	Y	O	10	D	0,0926	0,0934	0,0751	0,0793	0,0316	0,131
44	1000	Y	O	10	D	0,1167	0,1186	0,0902	0,0897	0,0004	0,1283
45	5000	Y	O	10	D	0,125	0,1239	0,1194	0,1184	0,0011	0,1397
46	250	D	Y	10	D	0,0871	0,0727	0,1241	0,1193	0,0627	0,1434
47	1000	D	Y	10	D	0,0681	0,0722	0,1023	0,141	0,037	0,14
48	5000	D	Y	10	D	0,0731	0,0871	0,0946	0,1793	0,1436	0,1306
49	250	O	Y	10	D	0,0535	0,0469	0,0918	0,0925	0,0298	0,0937
50	1000	O	Y	10	D	0,0281	0,0306	0,0668	0,0995	0,0049	0,0859
51	5000	O	Y	10	D	0,0137	0,0315	0,0499	0,1234	0,0368	0,0661
52	250	Y	Y	10	D	0,0387	0,0304	0,0765	0,0695	0,0239	0,0856
53	1000	Y	Y	10	D	0,013	0,0096	0,0429	0,0662	0,0002	0,0519
54	5000	Y	Y	10	D	0,0021	0,0015	0,0243	0,0717	0,0003	0,0238
55	250	D	D	25	D	0,0869	0,0623	0,0951	0,0954	0,036	0,1056
56	1000	D	D	25	D	0,0624	0,0468	0,0516	0,1074	0,0063	0,0961
57	5000	D	D	25	D	0,0291	0,0072	0,0107	0,1717	0,0134	0,1035
58	250	O	D	25	D	0,0645	0,0382	0,0768	0,0789	0,0169	0,0653
59	1000	O	D	25	D	0,0302	0,0148	0,0257	0,0809	0,0006	0,0566
60	5000	O	D	25	D	0,0028	0,0004	0,0016	0,1055	0	0,0491
61	250	Y	D	25	D	0,0391	0,0139	0,0662	0,0546	0,0024	0,0717
62	1000	Y	D	25	D	0,0127	0,0056	0,0232	0,0545	0,0004	0,0457
63	5000	Y	D	25	D	0,0008	0	0,0001	0,0591	0	0,0279
64	250	D	O	25	D	0,134	0,1393	0,1296	0,1753	0,0801	0,1932
65	1000	D	O	25	D	0,1463	0,1527	0,1575	0,2151	0,1197	0,2454
66	5000	D	O	25	D	0,1741	0,1696	0,1748	0,2538	0,0497	0,2618
67	250	O	O	25	D	0,1079	0,1133	0,0857	0,1193	0,0431	0,125
68	1000	O	O	25	D	0,1283	0,1268	0,1147	0,1462	0,0067	0,1546
69	5000	O	O	25	D	0,1445	0,1389	0,1433	0,1744	0,0365	0,1813
70	250	Y	O	25	D	0,0923	0,0981	0,0801	0,0822	0,0071	0,1145
71	1000	Y	O	25	D	0,1079	0,1151	0,0841	0,0949	0,0008	0,1218
72	5000	Y	O	25	D	0,124	0,1173	0,1157	0,1159	0,0006	0,1377
73	250	D	Y	25	D	0,1047	0,1035	0,1118	0,1468	0,0709	0,1432
74	1000	D	Y	25	D	0,1057	0,1014	0,1166	0,1833	0,0532	0,176
75	5000	D	Y	25	D	0,1151	0,1131	0,1116	0,2318	0,0618	0,1802
76	250	O	Y	25	D	0,0741	0,0699	0,0847	0,1	0,0294	0,1032
77	1000	O	Y	25	D	0,0559	0,0565	0,0821	0,1232	0,0026	0,1005
78	5000	O	Y	25	D	0,0368	0,0502	0,0584	0,1532	0	0,0997
79	250	Y	Y	25	D	0,0508	0,0448	0,0758	0,0748	0,0049	0,0818
80	1000	Y	Y	25	D	0,0266	0,0233	0,0465	0,0772	0,0004	0,0675
81	5000	Y	Y	25	D	0,0058	0,0067	0,023	0,0942	0	0,0581
82	250	D	D	5	O	0,2132	0,1381	0,2475	0,2631	0,1335	0,3155
83	1000	D	D	5	O	0,257	0,2134	0,2618	0,3378	0,1876	0,3691
84	5000	D	D	5	O	0,2974	0,2384	0,2593	0,3716	0,3995	0,3981
85	250	O	D	5	O	0,1161	0,0579	0,1872	0,1905	0,1338	0,1384

Tablo 4.2 – Devami

SNR	N	BSIZ_K	PRE	DS	BLI_K	C5.0	CHAID	C&RT	TAN	MB	DVM
86	1000	O	D	5	O	0,1003	0,0731	0,1586	0,2223	0,1139	0,0844
87	5000	O	D	5	O	0,1192	0,1121	0,1451	0,2577	0,3145	0,1405
88	250	Y	D	5	O	0,0535	0,0144	0,1342	0,1411	0,1316	0,0317
89	1000	Y	D	5	O	0,0284	0,0078	0,0894	0,137	0,0143	0,019
90	5000	Y	D	5	O	0,0129	0,0007	0,0509	0,1397	0,2422	0,0113
91	250	D	O	5	O	0,3749	0,3531	0,3773	0,4219	0,159	0,5123
92	1000	D	O	5	O	0,4281	0,4031	0,4263	0,4845	0,3929	0,5328
93	5000	D	O	5	O	0,4619	0,428	0,4469	0,5007	0,5054	0,539
94	250	O	O	5	O	0,3129	0,3105	0,2843	0,3276	0,3063	0,3903
95	1000	O	O	5	O	0,3386	0,3341	0,3356	0,3697	0,3626	0,4015
96	5000	O	O	5	O	0,3658	0,3572	0,367	0,388	0,3973	0,4101
97	250	Y	O	5	O	0,2775	0,2785	0,2162	0,2356	0,2928	0,3078
98	1000	Y	O	5	O	0,2917	0,2917	0,2741	0,289	0,1695	0,3198
99	5000	Y	O	5	O	0,305	0,307	0,3069	0,3026	0,3185	0,3267
100	250	D	Y	5	O	0,3364	0,301	0,3442	0,3701	0,164	0,4575
101	1000	D	Y	5	O	0,3786	0,3616	0,3778	0,4405	0,3056	0,4886
102	5000	D	Y	5	O	0,4172	0,377	0,3941	0,4638	0,4738	0,4961
103	250	O	Y	5	O	0,2313	0,2231	0,2347	0,2757	0,2513	0,3044
104	1000	O	Y	5	O	0,2762	0,2763	0,2813	0,3254	0,3072	0,3356
105	5000	O	Y	5	O	0,296	0,2941	0,3006	0,3476	0,3725	0,339
106	250	Y	Y	5	O	0,1382	0,161	0,1692	0,1956	0,2432	0,2013
107	1000	Y	Y	5	O	0,1823	0,2082	0,2048	0,2357	0,09	0,2162
108	5000	Y	Y	5	O	0,2153	0,2229	0,2257	0,2578	0,2982	0,2205
109	250	D	D	10	O	0,2932	0,2268	0,2787	0,3254	0,0927	0,4459
110	1000	D	D	10	O	0,3396	0,2806	0,2855	0,4621	0,0531	0,5182
111	5000	D	D	10	O	0,3771	0,2663	0,2818	0,522	0,3403	0,542
112	250	O	D	10	O	0,1646	0,0959	0,202	0,2147	0,0895	0,2116
113	1000	O	D	10	O	0,1692	0,1456	0,1942	0,2794	0,0588	0,2482
114	5000	O	D	10	O	0,1828	0,1651	0,1703	0,3262	0,3471	0,2339
115	250	Y	D	10	O	0,089	0,0352	0,1422	0,1488	0,1135	0,1285
116	1000	Y	D	10	O	0,0561	0,0138	0,0984	0,1699	0,0107	0,0865
117	5000	Y	D	10	O	0,0315	0,0015	0,0679	0,1987	0,2247	0,0556
118	250	D	O	10	O	0,4233	0,3771	0,4202	0,519	0,0181	0,6067
119	1000	D	O	10	O	0,4605	0,4178	0,4424	0,5839	0,0884	0,6346
120	5000	D	O	10	O	0,4995	0,4338	0,4564	0,6072	0,6146	0,6447
121	250	O	O	10	O	0,3118	0,309	0,2888	0,3476	0,185	0,3968
122	1000	O	O	10	O	0,3409	0,3324	0,3434	0,3955	0,1774	0,4206
123	5000	O	O	10	O	0,3723	0,3593	0,3713	0,416	0,4282	0,4339
124	250	Y	O	10	O	0,2644	0,2752	0,2001	0,2315	0,2873	0,3079
125	1000	Y	O	10	O	0,2918	0,2859	0,2644	0,2825	0,083	0,3179
126	5000	Y	O	10	O	0,3061	0,3054	0,305	0,2942	0,3255	0,3279
127	250	D	Y	10	O	0,3801	0,3307	0,3689	0,4569	0,0369	0,557
128	1000	D	Y	10	O	0,4258	0,3799	0,3996	0,5493	0,0514	0,602
129	5000	D	Y	10	O	0,4628	0,3887	0,4085	0,5835	0,5805	0,6156
130	250	O	Y	10	O	0,2574	0,2494	0,2604	0,3058	0,1649	0,3199
131	1000	O	Y	10	O	0,2847	0,2846	0,2952	0,3648	0,1689	0,3527
132	5000	O	Y	10	O	0,3143	0,3027	0,3118	0,3892	0,4035	0,3661
133	250	Y	Y	10	O	0,165	0,1686	0,1726	0,2007	0,2353	0,2081
134	1000	Y	Y	10	O	0,1977	0,2194	0,2041	0,2472	0,0457	0,2236
135	5000	Y	Y	10	O	0,2211	0,2283	0,2289	0,2673	0,3051	0,2246
136	250	D	D	25	O	0,3224	0,2824	0,2875	0,4028	0,0601	0,5473
137	1000	D	D	25	O	0,3722	0,3172	0,2916	0,5864	0,0529	0,6489
138	5000	D	D	25	O	0,4139	0,2836	0,293	0,6466	0,0096	0,6951
139	250	O	D	25	O	0,2029	0,1675	0,2126	0,243	0,077	0,2185
140	1000	O	D	25	O	0,2059	0,1975	0,2044	0,3232	0,0523	0,2578
141	5000	O	D	25	O	0,2126	0,187	0,1914	0,3507	0,3549	0,2774
142	250	Y	D	25	O	0,1211	0,0628	0,1433	0,1555	0,0835	0,1443
143	1000	Y	D	25	O	0,0927	0,0397	0,1115	0,1965	0,0056	0,1497
144	5000	Y	D	25	O	0,0604	0,0071	0,071	0,2295	0,1545	0,153
145	250	D	O	25	O	0,4486	0,4024	0,4324	0,6412	0,013	0,6862
146	1000	D	O	25	O	0,4898	0,4383	0,4495	0,6977	0	0,7418
147	5000	D	O	25	O	0,5264	0,4341	0,4596	0,7303	0,7384	0,7612
148	250	O	O	25	O	0,2915	0,2962	0,2961	0,3949	0,0441	0,381
149	1000	O	O	25	O	0,3297	0,3342	0,3529	0,4299	0,01	0,429
150	5000	O	O	25	O	0,3659	0,3593	0,3677	0,4375	0,4486	0,4464
151	250	Y	O	25	O	0,2511	0,2659	0,1922	0,2367	0,2777	0,2935
152	1000	Y	O	25	O	0,2832	0,2789	0,2589	0,2897	0,0383	0,3202
153	5000	Y	O	25	O	0,3017	0,2994	0,3009	0,2877	0,3299	0,3268
154	250	D	Y	25	O	0,4083	0,3671	0,3914	0,5933	0,0297	0,6588
155	1000	D	Y	25	O	0,4568	0,4072	0,4079	0,6707	0	0,7217
156	5000	D	Y	25	O	0,4939	0,3927	0,4144	0,7091	0,4891	0,7429
157	250	O	Y	25	O	0,2584	0,2534	0,2613	0,3516	0,0551	0,3056
158	1000	O	Y	25	O	0,2882	0,2941	0,3093	0,4027	0,0344	0,3591
159	5000	O	Y	25	O	0,3189	0,314	0,3158	0,4117	0,4216	0,3725

Tablo 4.2 – Devamı

SNR	N	BSIZ_K	PRE	DS	BLI_K	C5.0	CHAID	C&RT	TAN	MB	DVM
160	250	Y	Y	25	O	0,1894	0,1821	0,1805	0,2112	0,2136	0,2179
161	1000	Y	Y	25	O	0,2076	0,2197	0,2083	0,2606	0,0293	0,218
162	5000	Y	Y	25	O	0,2239	0,228	0,2296	0,267	0,3073	0,2193
169	250	Y	D	5	Y	0,4667	0,4487	0,4651	0,5014	0,5633	0,5165
170	1000	Y	D	5	Y	0,5121	0,5209	0,5216	0,5516	0,5594	0,5737
171	5000	Y	D	5	Y	0,5421	0,5255	0,5365	0,5588	0,5698	0,5869
178	250	Y	O	5	Y	0,5946	0,5938	0,5781	0,6168	0,6592	0,6703
179	1000	Y	O	5	Y	0,6257	0,622	0,6264	0,6342	0,6545	0,6712
180	5000	Y	O	5	Y	0,6451	0,6352	0,6438	0,6315	0,6475	0,6747
187	250	Y	Y	5	Y	0,5627	0,5621	0,5493	0,58	0,6334	0,6244
188	1000	Y	Y	5	Y	0,5934	0,5959	0,5959	0,6191	0,6366	0,6455
189	5000	Y	Y	5	Y	0,617	0,6107	0,6166	0,623	0,6377	0,6499
196	250	Y	D	10	Y	0,4816	0,4916	0,4883	0,5063	0,5668	0,5124
197	1000	Y	D	10	Y	0,5271	0,5225	0,5255	0,5629	0,5433	0,5876
198	5000	Y	D	10	Y	0,5512	0,5362	0,539	0,5666	0,5724	0,6035
205	250	Y	O	10	Y	0,6015	0,5985	0,586	0,6225	0,6818	0,6731
206	1000	Y	O	10	Y	0,629	0,6268	0,6316	0,6496	0,6811	0,6858
207	5000	Y	O	10	Y	0,6483	0,6383	0,6461	0,6426	0,6732	0,6871
214	250	Y	Y	10	Y	0,5608	0,5676	0,5617	0,5956	0,6506	0,6319
215	1000	Y	Y	10	Y	0,5961	0,5976	0,6026	0,6324	0,6519	0,6546
216	5000	Y	Y	10	Y	0,6195	0,6126	0,618	0,6312	0,6528	0,6603
223	250	Y	D	25	Y	0,491	0,5102	0,5057	0,5308	0,5691	0,5293
224	1000	Y	D	25	Y	0,5146	0,5299	0,5333	0,5742	0,5034	0,5796
225	5000	Y	D	25	Y	0,5518	0,5441	0,545	0,5599	0,5662	0,6037
232	250	Y	O	25	Y	0,5857	0,5937	0,5945	0,6226	0,6901	0,6635
233	1000	Y	O	25	Y	0,6116	0,6258	0,6331	0,6639	0,6953	0,6859
234	5000	Y	O	25	Y	0,6438	0,6371	0,6468	0,6574	0,6906	0,6926
241	250	Y	Y	25	Y	0,5599	0,5667	0,5713	0,5959	0,6585	0,6258
242	1000	Y	Y	25	Y	0,5807	0,5973	0,606	0,6449	0,6611	0,6555
243	5000	Y	Y	25	Y	0,6196	0,6146	0,6206	0,6371	0,6629	0,6654

Tablo 4.3. Yöntemlerin Senaryo Bazlı F Kriterine Göre Sonuçları

SNR	N	BSIZ_K	PRE	DS	BLI_K	C5.0	CHAID	C&RT	TAN	MB	DVM
1	250	D	D	5	D	0,0093	0,0053	0,0784	0,0853	0,0129	0,0207
2	1000	D	D	5	D	0,0048	0,008	0,0281	0,0585	0,0016	0,0026
3	5000	D	D	5	D	0,0004	0,0001	0,001	0,027	0,0333	0,0001
4	250	O	D	5	D	0,0106	0,006	0,062	0,0832	0,0231	0,0077
5	1000	O	D	5	D	0,002	0,0013	0,0162	0,0453	0,0034	0,0018
6	5000	O	D	5	D	0	0	0,0002	0,0197	0,0055	0
7	250	Y	D	5	D	0,0058	0,0039	0,0465	0,0492	0,0151	0,0024
8	1000	Y	D	5	D	0,0011	0,0006	0,0103	0,0311	0,0004	0,0008
9	5000	Y	D	5	D	0	0	0	0,0101	0	0,0002
10	250	D	O	5	D	0,4957	0,476	0,5271	0,5463	0,4505	0,1943
11	1000	D	O	5	D	0,55	0,5506	0,5611	0,5742	0,5132	0,2058
12	5000	D	O	5	D	0,5767	0,5777	0,5791	0,5967	0,6007	0,2141
13	250	O	O	5	D	0,468	0,4213	0,5139	0,5338	0,4336	0,1496
14	1000	O	O	5	D	0,5355	0,5459	0,5437	0,5578	0,3744	0,1623
15	5000	O	O	5	D	0,5628	0,5668	0,5665	0,5762	0,5604	0,1688
16	250	Y	O	5	D	0,4354	0,3452	0,5046	0,5182	0,3963	0,1332
17	1000	Y	O	5	D	0,5044	0,5339	0,5226	0,543	0,3287	0,1328
18	5000	Y	O	5	D	0,547	0,5512	0,5527	0,5576	0,3473	0,1361
19	250	D	Y	5	D	0,8425	0,6771	0,7997	0,8044	0,8396	0,0954
20	1000	D	Y	5	D	0,8531	0,8502	0,8376	0,8388	0,8532	0,0681
21	5000	D	Y	5	D	0,8557	0,8529	0,8516	0,8501	0,8438	0,0693
22	250	O	Y	5	D	0,8469	0,6243	0,8034	0,799	0,8095	0,0724
23	1000	O	Y	5	D	0,8554	0,8519	0,8395	0,8383	0,8531	0,0359
24	5000	O	Y	5	D	0,8572	0,8562	0,8542	0,8501	0,8394	0,0252
25	250	Y	Y	5	D	0,8495	0,576	0,8063	0,8039	0,8311	0,05
26	1000	Y	Y	5	D	0,8561	0,8485	0,8425	0,8434	0,8565	0,019
27	5000	Y	Y	5	D	0,857	0,857	0,8553	0,853	0,8568	0,0077
28	250	D	D	10	D	0,0543	0,0297	0,1172	0,1146	0,0226	0,0978
29	1000	D	D	10	D	0,0269	0,0153	0,0481	0,1322	0,0033	0,0445
30	5000	D	D	10	D	0,0051	0,0013	0,0044	0,1242	0,0302	0,0087
31	250	O	D	10	D	0,038	0,0118	0,0773	0,1052	0,0226	0,0793
32	1000	O	D	10	D	0,0105	0,0043	0,026	0,1	0,0029	0,0336
33	5000	O	D	10	D	0,0003	0,0001	0,0005	0,0904	0,001	0,0005
34	250	Y	D	10	D	0,0228	0,0059	0,0676	0,0916	0,019	0,0372
35	1000	Y	D	10	D	0,0072	0,0008	0,0158	0,0627	0,0007	0,0017
36	5000	Y	D	10	D	0	0	0,0001	0,0426	0	0
37	250	D	O	10	D	0,5229	0,5151	0,5428	0,5615	0,4527	0,201
38	1000	D	O	10	D	0,5682	0,5602	0,5721	0,5888	0,5255	0,2331
39	5000	D	O	10	D	0,5831	0,5808	0,5841	0,6113	0,544	0,241

Tablo 4.3 – Devami

SNR	N	BSIZ_K	PRE	DS	BLIK	C5.0	CHAID	C&RT	TAN	MB	DVM
40	250	O	O	10	D	0,4891	0,4695	0,5201	0,5325	0,3928	0,1402
41	1000	O	O	10	D	0,5422	0,5472	0,544	0,5625	0,3617	0,1598
42	5000	O	O	10	D	0,5659	0,5642	0,566	0,582	0,5097	0,176
43	250	Y	O	10	D	0,4517	0,3957	0,5022	0,5126	0,3859	0,131
44	1000	Y	O	10	D	0,5181	0,5388	0,5254	0,5399	0,3183	0,1283
45	5000	Y	O	10	D	0,5412	0,551	0,5484	0,5579	0,3364	0,1397
46	250	D	Y	10	D	0,8365	0,7909	0,8058	0,795	0,8111	0,1434
47	1000	D	Y	10	D	0,8424	0,8396	0,8318	0,8178	0,8493	0,14
48	5000	D	Y	10	D	0,8513	0,8489	0,8485	0,8337	0,8079	0,1306
49	250	O	Y	10	D	0,8406	0,7245	0,8001	0,784	0,8283	0,0937
50	1000	O	Y	10	D	0,8523	0,8495	0,8366	0,8134	0,851	0,0859
51	5000	O	Y	10	D	0,8565	0,8546	0,8527	0,8273	0,8322	0,0661
52	250	Y	Y	10	D	0,8446	0,6108	0,8001	0,7761	0,8273	0,0856
53	1000	Y	Y	10	D	0,8548	0,8501	0,8378	0,8215	0,8563	0,0519
54	5000	Y	Y	10	D	0,8574	0,8572	0,8549	0,8376	0,8573	0,0238
55	250	D	D	25	D	0,1113	0,0782	0,117	0,1128	0,0494	0,1056
56	1000	D	D	25	D	0,0859	0,0542	0,0588	0,1773	0,0058	0,0961
57	5000	D	D	25	D	0,0275	0,0064	0,0092	0,2605	0,0133	0,1035
58	250	O	D	25	D	0,0795	0,0449	0,0916	0,0996	0,0297	0,0653
59	1000	O	D	25	D	0,0356	0,0165	0,0294	0,1545	0,0014	0,0566
60	5000	O	D	25	D	0,0025	0,0004	0,0014	0,2055	0	0,0491
61	250	Y	D	25	D	0,0472	0,0181	0,0816	0,0918	0,0046	0,0717
62	1000	Y	D	25	D	0,0135	0,0055	0,0255	0,1267	0,0008	0,0457
63	5000	Y	D	25	D	0,0005	0	0,0002	0,1461	0	0,0279
64	250	D	O	25	D	0,5384	0,5365	0,5469	0,5784	0,3709	0,1932
65	1000	D	O	25	D	0,5608	0,5649	0,5736	0,6062	0,5299	0,2454
66	5000	D	O	25	D	0,5819	0,5815	0,5822	0,6265	0,2833	0,2618
67	250	O	O	25	D	0,504	0,5089	0,5183	0,5475	0,4015	0,125
68	1000	O	O	25	D	0,5377	0,5484	0,5487	0,5709	0,3272	0,1546
69	5000	O	O	25	D	0,5673	0,565	0,5653	0,5867	0,3938	0,1813
70	250	Y	O	25	D	0,4533	0,4454	0,5103	0,5179	0,3219	0,1145
71	1000	Y	O	25	D	0,5128	0,5282	0,5218	0,5439	0,3008	0,1218
72	5000	Y	O	25	D	0,5496	0,5455	0,5476	0,5565	0,3336	0,1377
73	250	D	Y	25	D	0,7963	0,7971	0,7911	0,8028	0,6534	0,1432
74	1000	D	Y	25	D	0,8234	0,8239	0,8268	0,7986	0,8376	0,176
75	5000	D	Y	25	D	0,8414	0,8425	0,8449	0,8003	0,5924	0,1802
76	250	O	Y	25	D	0,8186	0,7738	0,7905	0,7817	0,8133	0,1032
77	1000	O	Y	25	D	0,8431	0,8375	0,8311	0,7815	0,8531	0,1005
78	5000	O	Y	25	D	0,8535	0,8503	0,8496	0,7758	0,8561	0,0997
79	250	Y	Y	25	D	0,8367	0,7034	0,7894	0,7574	0,8473	0,0818
80	1000	Y	Y	25	D	0,8513	0,8472	0,83	0,7778	0,857	0,0675
81	5000	Y	Y	25	D	0,8566	0,8561	0,8542	0,7875	0,8568	0,0581
82	250	D	D	5	O	0,231	0,1501	0,2743	0,2953	0,136	0,3155
83	1000	D	D	5	O	0,2788	0,2271	0,2847	0,3692	0,1918	0,3691
84	5000	D	D	5	O	0,3074	0,2328	0,2612	0,3976	0,442	0,3981
85	250	O	D	5	O	0,1285	0,0658	0,2137	0,2259	0,1863	0,1384
86	1000	O	D	5	O	0,101	0,0788	0,1673	0,258	0,1582	0,0844
87	5000	O	D	5	O	0,1038	0,1026	0,1346	0,2898	0,3857	0,1405
88	250	Y	D	5	O	0,0578	0,0175	0,1548	0,1646	0,1723	0,0317
89	1000	Y	D	5	O	0,028	0,0098	0,0913	0,1495	0,0143	0,019
90	5000	Y	D	5	O	0,0105	0,0007	0,0441	0,1354	0,3212	0,0113
91	250	D	O	5	O	0,6804	0,6614	0,6826	0,7044	0,3398	0,5123
92	1000	D	O	5	O	0,7116	0,6975	0,7098	0,7412	0,6644	0,5328
93	5000	D	O	5	O	0,7305	0,7127	0,7229	0,7502	0,7521	0,539
94	250	O	O	5	O	0,6378	0,633	0,6307	0,6563	0,6389	0,3903
95	1000	O	O	5	O	0,6629	0,662	0,6632	0,683	0,6772	0,4015
96	5000	O	O	5	O	0,6819	0,6765	0,6823	0,6931	0,6975	0,4101
97	250	Y	O	5	O	0,6085	0,6099	0,5942	0,6081	0,6392	0,3078
98	1000	Y	O	5	O	0,6349	0,6343	0,6289	0,6407	0,5725	0,3198
99	5000	Y	O	5	O	0,6508	0,651	0,6519	0,6475	0,6561	0,3267
100	250	D	Y	5	O	0,8441	0,834	0,8442	0,8509	0,5375	0,4575
101	1000	D	Y	5	O	0,8596	0,8548	0,8594	0,8722	0,7882	0,4886
102	5000	D	Y	5	O	0,8706	0,8642	0,8678	0,8783	0,8767	0,4961
103	250	O	Y	5	O	0,8379	0,8311	0,823	0,8301	0,8005	0,3044
104	1000	O	Y	5	O	0,8514	0,849	0,8489	0,8492	0,8333	0,3356
105	5000	O	Y	5	O	0,8594	0,857	0,8585	0,8546	0,8391	0,339
106	250	Y	Y	5	O	0,8436	0,8371	0,815	0,8212	0,7979	0,2013
107	1000	Y	Y	5	O	0,853	0,8513	0,8464	0,8425	0,8463	0,2162
108	5000	Y	Y	5	O	0,8559	0,8562	0,8549	0,8451	0,811	0,2205
109	250	D	D	10	O	0,327	0,2504	0,3096	0,362	0,1079	0,4459
110	1000	D	D	10	O	0,3888	0,3153	0,3181	0,5081	0,0502	0,5182
111	5000	D	D	10	O	0,4187	0,2776	0,2955	0,5689	0,3596	0,542
112	250	O	D	10	O	0,1846	0,1077	0,2325	0,2644	0,1581	0,2116
113	1000	O	D	10	O	0,1857	0,1586	0,2135	0,3486	0,1627	0,2482

Tablo 4.3 – Devami

SNR	N	BSIZ_K	PRE	DS	BLIK	C5.0	CHAID	C&RT	TAN	MB	DVM
114	5000	O	D	10	O	0,1729	0,1489	0,1628	0,3969	0,4097	0,2339
115	250	Y	D	10	O	0,0998	0,04	0,1656	0,196	0,1682	0,1285
116	1000	Y	D	10	O	0,0547	0,0165	0,1024	0,2362	0,0218	0,0865
117	5000	Y	D	10	O	0,0253	0,0016	0,0601	0,2725	0,2915	0,0556
118	250	D	O	10	O	0,7098	0,6788	0,7047	0,7551	0,032	0,6067
119	1000	D	O	10	O	0,7291	0,7047	0,719	0,7906	0,1339	0,6346
120	5000	D	O	10	O	0,7498	0,7155	0,7274	0,8036	0,8071	0,6447
121	250	O	O	10	O	0,6422	0,6377	0,636	0,6683	0,4816	0,3968
122	1000	O	O	10	O	0,6668	0,6614	0,6678	0,6966	0,5151	0,4206
123	5000	O	O	10	O	0,6843	0,6772	0,6838	0,7078	0,7137	0,4339
124	250	Y	O	10	O	0,5994	0,6097	0,5855	0,6092	0,6342	0,3079
125	1000	Y	O	10	O	0,6327	0,6314	0,6257	0,6381	0,5312	0,3179
126	5000	Y	O	10	O	0,6503	0,6498	0,6491	0,6469	0,6623	0,3279
127	250	D	Y	10	O	0,8478	0,8322	0,8478	0,867	0,0879	0,557
128	1000	D	Y	10	O	0,8615	0,8529	0,861	0,8883	0,0599	0,602
129	5000	D	Y	10	O	0,8733	0,8631	0,8677	0,8962	0,8929	0,6156
130	250	O	Y	10	O	0,8328	0,8278	0,8251	0,8246	0,6367	0,3199
131	1000	O	Y	10	O	0,8471	0,8455	0,8468	0,8377	0,7021	0,3527
132	5000	O	Y	10	O	0,8574	0,8556	0,8573	0,8421	0,8274	0,3661
133	250	Y	Y	10	O	0,8395	0,8319	0,8119	0,8016	0,7777	0,2081
134	1000	Y	Y	10	O	0,8508	0,8489	0,8434	0,8199	0,8087	0,2236
135	5000	Y	Y	10	O	0,8557	0,8556	0,8544	0,822	0,7907	0,2246
136	250	D	D	25	O	0,3622	0,3229	0,3207	0,4281	0,0936	0,5473
137	1000	D	D	25	O	0,426	0,3636	0,3277	0,6244	0,0767	0,6489
138	5000	D	D	25	O	0,4645	0,3073	0,3168	0,6738	0,0098	0,6951
139	250	O	D	25	O	0,2406	0,1937	0,2451	0,2935	0,1381	0,2185
140	1000	O	D	25	O	0,2514	0,2229	0,2322	0,3922	0,1409	0,2578
141	5000	O	D	25	O	0,2264	0,1722	0,1886	0,4092	0,4019	0,2774
142	250	Y	D	25	O	0,1422	0,0744	0,1727	0,2129	0,1188	0,1443
143	1000	Y	D	25	O	0,1015	0,0455	0,123	0,2843	0,0102	0,1497
144	5000	Y	D	25	O	0,0492	0,0076	0,0635	0,3145	0,208	0,153
145	250	D	O	25	O	0,7215	0,6934	0,7103	0,8152	0,0438	0,6862
146	1000	D	O	25	O	0,7446	0,7156	0,7229	0,8483	0	0,7418
147	5000	D	O	25	O	0,7631	0,7149	0,7288	0,865	0,869	0,7612
148	250	O	O	25	O	0,6405	0,6359	0,642	0,6936	0,138	0,381
149	1000	O	O	25	O	0,6604	0,6615	0,6734	0,7138	0,0673	0,429
150	5000	O	O	25	O	0,6816	0,6775	0,6826	0,7183	0,724	0,4464
151	250	Y	O	25	O	0,5926	0,6107	0,5838	0,6094	0,6241	0,2935
152	1000	Y	O	25	O	0,6274	0,6273	0,6236	0,643	0,4415	0,3202
153	5000	Y	O	25	O	0,6461	0,6463	0,6473	0,6429	0,6639	0,3268
154	250	D	Y	25	O	0,8495	0,833	0,8508	0,899	0,1091	0,6588
155	1000	D	Y	25	O	0,8654	0,8546	0,8608	0,914	0	0,7217
156	5000	D	Y	25	O	0,8769	0,8624	0,8672	0,9209	0,8252	0,7429
157	250	O	Y	25	O	0,8203	0,8144	0,8216	0,8266	0,2693	0,3056
158	1000	O	Y	25	O	0,8367	0,8399	0,8451	0,8288	0,2532	0,3591
159	5000	O	Y	25	O	0,8534	0,8541	0,8568	0,8247	0,8168	0,3725
160	250	Y	Y	25	O	0,83	0,823	0,8082	0,7831	0,7747	0,2179
161	1000	Y	Y	25	O	0,8457	0,843	0,839	0,7905	0,8301	0,218
162	5000	Y	Y	25	O	0,8549	0,854	0,8533	0,7848	0,7733	0,2193
169	250	Y	D	5	Y	0,4849	0,4677	0,4895	0,5311	0,5843	0,5165
170	1000	Y	D	5	Y	0,5333	0,5403	0,5468	0,5893	0,585	0,5737
171	5000	Y	D	5	Y	0,5674	0,5456	0,5618	0,6009	0,5977	0,5869
178	250	Y	O	5	Y	0,7909	0,7888	0,7839	0,8042	0,826	0,6703
179	1000	Y	O	5	Y	0,8098	0,8079	0,8105	0,8147	0,8249	0,6712
180	5000	Y	O	5	Y	0,8218	0,817	0,8211	0,8124	0,8206	0,6747
187	250	Y	Y	5	Y	0,892	0,8912	0,8886	0,8938	0,8953	0,6244
188	1000	Y	Y	5	Y	0,9031	0,9026	0,9025	0,9041	0,8974	0,6455
189	5000	Y	Y	5	Y	0,909	0,9069	0,9087	0,9056	0,9004	0,6499
196	250	Y	D	10	Y	0,5056	0,5105	0,5135	0,5401	0,5794	0,5124
197	1000	Y	D	10	Y	0,5571	0,5459	0,554	0,5994	0,5664	0,5876
198	5000	Y	D	10	Y	0,5797	0,5598	0,566	0,6034	0,5906	0,6035
205	250	Y	O	10	Y	0,7945	0,7895	0,7872	0,8067	0,8367	0,6731
206	1000	Y	O	10	Y	0,8126	0,8109	0,8144	0,8242	0,8398	0,6858
207	5000	Y	O	10	Y	0,8239	0,8185	0,8225	0,8209	0,8361	0,6871
214	250	Y	Y	10	Y	0,8907	0,891	0,8909	0,8913	0,8955	0,6319
215	1000	Y	Y	10	Y	0,9022	0,9017	0,9038	0,902	0,8974	0,6546
216	5000	Y	Y	10	Y	0,9092	0,907	0,9089	0,902	0,9004	0,6603
223	250	Y	D	25	Y	0,5202	0,5313	0,5321	0,5618	0,5784	0,5293
224	1000	Y	D	25	Y	0,5532	0,5585	0,565	0,6031	0,5399	0,5796
225	5000	Y	D	25	Y	0,5845	0,5725	0,5747	0,587	0,5788	0,6037
232	250	Y	O	25	Y	0,7896	0,7901	0,7938	0,8085	0,842	0,6635
233	1000	Y	O	25	Y	0,8046	0,8105	0,8147	0,8313	0,847	0,6859
234	5000	Y	O	25	Y	0,8213	0,8174	0,8225	0,8283	0,8449	0,6926
241	250	Y	Y	25	Y	0,8891	0,8894	0,8928	0,8865	0,8953	0,6258

Tablo 4.3 – Devamı

SNR	N	BSIZ_K	PRE	DS	BLL_K	C5.0	CHAID	C&RT	TAN	MB	DVM
242	1000	Y	Y	25	Y	0,896	0,9007	0,904	0,8996	0,8969	0,6555
243	5000	Y	Y	25	Y	0,9083	0,9071	0,9093	0,8978	0,8998	0,6654

Tablo 4.4. Yöntemlerin Senaryo Bazlı GD Kriterine Göre Sonuçları

SNR	N	BSIZ_K	PRE	DS	BLL_K	C5.0	CHAID	C&RT	TAN	MB	DVM
1	250	D	D	5	D	0,8927	0,5506	0,7698	0,8602	0,8946	0,8927
2	1000	D	D	5	D	0,8807	0,8665	0,826	0,8892	0,8984	0,8981
3	5000	D	D	5	D	0,9017	0,9017	0,8998	0,8994	0,8982	0,9016
4	250	O	D	5	D	0,8989	0,5362	0,748	0,8617	0,8677	0,8986
5	1000	O	D	5	D	0,9015	0,8854	0,8014	0,8914	0,8998	0,9014
6	5000	O	D	5	D	0,9015	0,9015	0,896	0,8991	0,9004	0,9014
7	250	Y	D	5	D	0,8978	0,4519	0,7293	0,8663	0,8819	0,8983
8	1000	Y	D	5	D	0,9019	0,8463	0,7916	0,894	0,9019	0,9019
9	5000	Y	D	5	D	0,9015	0,9015	0,8925	0,8999	0,9015	0,9014
10	250	D	O	5	D	0,5434	0,4953	0,5449	0,5537	0,5181	0,5918
11	1000	D	O	5	D	0,5631	0,5662	0,5678	0,5758	0,5451	0,6017
12	5000	D	O	5	D	0,5804	0,5805	0,5828	0,5971	0,6001	0,6067
13	250	O	O	5	D	0,528	0,44	0,5234	0,5377	0,4925	0,562
14	1000	O	O	5	D	0,5559	0,5589	0,5532	0,5593	0,4962	0,5792
15	5000	O	O	5	D	0,5664	0,5692	0,5686	0,5774	0,5704	0,5837
16	250	Y	O	5	D	0,5211	0,3673	0,5201	0,5219	0,4917	0,5494
17	1000	Y	O	5	D	0,5458	0,5523	0,5398	0,5461	0,4906	0,5632
18	5000	Y	O	5	D	0,558	0,5606	0,5586	0,56	0,4982	0,5668
19	250	D	Y	5	D	0,7351	0,5896	0,6863	0,6923	0,7306	0,7236
20	1000	D	Y	5	D	0,7465	0,7434	0,73	0,7322	0,7462	0,7485
21	5000	D	Y	5	D	0,7489	0,7465	0,7456	0,7457	0,7407	0,7508
22	250	O	Y	5	D	0,7395	0,5441	0,6882	0,682	0,696	0,7237
23	1000	O	Y	5	D	0,7486	0,745	0,7308	0,7295	0,7454	0,7482
24	5000	O	Y	5	D	0,7503	0,7493	0,7475	0,7437	0,7326	0,7502
25	250	Y	Y	5	D	0,7425	0,5026	0,6909	0,6858	0,7196	0,7336
26	1000	Y	Y	5	D	0,7492	0,7422	0,7337	0,7344	0,7495	0,7481
27	5000	Y	Y	5	D	0,7499	0,7499	0,7482	0,7458	0,7497	0,7497
28	250	D	D	10	D	0,8858	0,7486	0,8189	0,8533	0,8899	0,8369
29	1000	D	D	10	D	0,8974	0,8986	0,806	0,8751	0,9012	0,8967
30	5000	D	D	10	D	0,9011	0,9015	0,8931	0,8881	0,8951	0,9016
31	250	O	D	10	D	0,8926	0,6513	0,798	0,848	0,8796	0,8451
32	1000	O	D	10	D	0,9007	0,8982	0,8022	0,8739	0,9	0,8995
33	5000	O	D	10	D	0,9014	0,9014	0,8951	0,8862	0,9006	0,9014
34	250	Y	D	10	D	0,8974	0,5049	0,7685	0,8444	0,8791	0,8873
35	1000	Y	D	10	D	0,9008	0,8764	0,761	0,8774	0,9013	0,9014
36	5000	Y	D	10	D	0,9014	0,9014	0,8951	0,8921	0,9014	0,9014
37	250	D	O	10	D	0,55	0,5385	0,5526	0,5656	0,5239	0,5955
38	1000	D	O	10	D	0,5743	0,5713	0,5757	0,591	0,5538	0,6156
39	5000	D	O	10	D	0,5857	0,5825	0,5857	0,6111	0,5858	0,6201
40	250	O	O	10	D	0,5359	0,5022	0,5312	0,5375	0,4954	0,5603
41	1000	O	O	10	D	0,5568	0,5579	0,552	0,5629	0,4927	0,5778
42	5000	O	O	10	D	0,569	0,5693	0,5697	0,5818	0,5479	0,5871
43	250	Y	O	10	D	0,5224	0,417	0,5179	0,5186	0,4877	0,5515
44	1000	Y	O	10	D	0,5489	0,5537	0,5387	0,5418	0,4906	0,5611
45	5000	Y	O	10	D	0,5593	0,5602	0,558	0,5589	0,4965	0,5693
46	250	D	Y	10	D	0,7312	0,6877	0,6969	0,6844	0,7072	0,7055
47	1000	D	Y	10	D	0,7347	0,7315	0,7241	0,7118	0,7414	0,7399
48	5000	D	Y	10	D	0,7453	0,7429	0,743	0,7328	0,7049	0,753
49	250	O	Y	10	D	0,7329	0,6294	0,6865	0,6693	0,7177	0,6903
50	1000	O	Y	10	D	0,7452	0,7417	0,7269	0,7029	0,7429	0,7374
51	5000	O	Y	10	D	0,7497	0,7478	0,746	0,7206	0,7233	0,7492
52	250	Y	Y	10	D	0,7367	0,5307	0,6839	0,6576	0,7157	0,7099
53	1000	Y	Y	10	D	0,7477	0,7429	0,7277	0,7087	0,7493	0,7424
54	5000	Y	Y	10	D	0,7505	0,7503	0,7479	0,7285	0,7504	0,7497
55	250	D	D	25	D	0,859	0,8426	0,8546	0,8655	0,8679	0,8224
56	1000	D	D	25	D	0,8811	0,8899	0,8014	0,8535	0,8996	0,826
57	5000	D	D	25	D	0,8985	0,9014	0,894	0,8343	0,8978	0,8608
58	250	O	D	25	D	0,8741	0,8139	0,8121	0,8546	0,8709	0,8133
59	1000	O	D	25	D	0,8931	0,897	0,7579	0,8411	0,9005	0,8136
60	5000	O	D	25	D	0,9011	0,8998	0,8943	0,8114	0,9036	0,8653
61	250	Y	D	25	D	0,8864	0,6464	0,7565	0,8268	0,8965	0,8369
62	1000	Y	D	25	D	0,8959	0,884	0,7555	0,833	0,8997	0,8367
63	5000	Y	D	25	D	0,9018	0,8983	0,8892	0,8349	0,9017	0,897
64	250	D	O	25	D	0,5547	0,5558	0,5551	0,5828	0,5211	0,5918
65	1000	D	O	25	D	0,571	0,5746	0,5771	0,6073	0,5558	0,6218
66	5000	D	O	25	D	0,5861	0,5843	0,5864	0,6268	0,5223	0,6304
67	250	O	O	25	D	0,5342	0,5303	0,5244	0,5496	0,4952	0,552

Tablo 4.4 – Devami

SNR	N	BSIZ_K	PRE	DS	BLIK	C5.0	CHAID	C&RT	TAN	MB	DVM
68	1000	O	O	25	D	0,5589	0,5594	0,5538	0,5724	0,4909	0,5758
69	5000	O	O	25	D	0,5706	0,5687	0,5703	0,5871	0,5093	0,5902
70	250	Y	O	25	D	0,5191	0,4826	0,5151	0,5231	0,4826	0,5431
71	1000	Y	O	25	D	0,5416	0,5515	0,5362	0,5444	0,4894	0,5585
72	5000	Y	O	25	D	0,5588	0,5568	0,556	0,5577	0,4946	0,5683
73	250	D	Y	25	D	0,6855	0,6875	0,6797	0,6963	0,5879	0,6874
74	1000	D	Y	25	D	0,7152	0,7153	0,7196	0,6969	0,7296	0,7282
75	5000	D	Y	25	D	0,736	0,737	0,7398	0,7044	0,5415	0,7541
76	250	O	Y	25	D	0,7077	0,6681	0,6754	0,6683	0,7012	0,6721
77	1000	O	Y	25	D	0,7346	0,728	0,7216	0,6722	0,7453	0,7032
78	5000	O	Y	25	D	0,7466	0,7429	0,7425	0,6703	0,7483	0,7383
79	250	Y	Y	25	D	0,7278	0,6098	0,6722	0,6392	0,7386	0,6786
80	1000	Y	Y	25	D	0,7438	0,7386	0,7194	0,6627	0,7503	0,7061
81	5000	Y	Y	25	D	0,7495	0,749	0,7468	0,6749	0,7496	0,7367
82	250	D	D	5	O	0,8847	0,8858	0,8796	0,8822	0,8887	0,9018
83	1000	D	D	5	O	0,8964	0,8958	0,8956	0,9019	0,8955	0,9111
84	5000	D	D	5	O	0,9035	0,9017	0,9018	0,908	0,9047	0,9145
85	250	O	D	5	O	0,891	0,8851	0,8733	0,8677	0,8135	0,8942
86	1000	O	D	5	O	0,8972	0,8967	0,8903	0,8877	0,8692	0,9014
87	5000	O	D	5	O	0,9007	0,9005	0,9001	0,8947	0,8644	0,9034
88	250	Y	D	5	O	0,8947	0,874	0,8636	0,8695	0,8351	0,898
89	1000	Y	D	5	O	0,9011	0,9016	0,8945	0,8909	0,8996	0,9026
90	5000	Y	D	5	O	0,9008	0,901	0,9001	0,8954	0,8196	0,901
91	250	D	O	5	O	0,683	0,6714	0,6866	0,7093	0,5581	0,7548
92	1000	D	O	5	O	0,7125	0,7002	0,712	0,7417	0,6929	0,766
93	5000	D	O	5	O	0,7305	0,7136	0,7228	0,75	0,752	0,7694
94	250	O	O	5	O	0,6485	0,6473	0,6378	0,6618	0,6498	0,693
95	1000	O	O	5	O	0,6663	0,6649	0,6659	0,6843	0,681	0,7003
96	5000	O	O	5	O	0,6821	0,6781	0,6827	0,6937	0,6981	0,7049
97	250	Y	O	5	O	0,627	0,6286	0,601	0,6135	0,6437	0,6504
98	1000	Y	O	5	O	0,6403	0,6417	0,6338	0,6429	0,5809	0,6588
99	5000	Y	O	5	O	0,651	0,6525	0,6521	0,6498	0,6582	0,6629
100	250	D	Y	5	O	0,7622	0,748	0,7631	0,7732	0,5044	0,8097
101	1000	D	Y	5	O	0,7821	0,7754	0,7818	0,8023	0,7052	0,8211
102	5000	D	Y	5	O	0,7975	0,7863	0,7919	0,8111	0,8108	0,8244
103	250	O	Y	5	O	0,7444	0,7372	0,7283	0,7403	0,7088	0,7687
104	1000	O	Y	5	O	0,762	0,7596	0,7599	0,7654	0,7473	0,7829
105	5000	O	Y	5	O	0,7718	0,7693	0,7715	0,7732	0,7607	0,7855
106	250	Y	Y	5	O	0,7435	0,7384	0,7119	0,7214	0,7045	0,7526
107	1000	Y	Y	5	O	0,756	0,756	0,7495	0,7485	0,7408	0,7635
108	5000	Y	Y	5	O	0,7602	0,761	0,7601	0,7531	0,7236	0,7654
109	250	D	D	10	O	0,8829	0,8815	0,8843	0,8878	0,8613	0,9129
110	1000	D	D	10	O	0,8922	0,8925	0,8943	0,9066	0,8851	0,9244
111	5000	D	D	10	O	0,9031	0,9005	0,9008	0,9136	0,9019	0,9284
112	250	O	D	10	O	0,8812	0,8858	0,8739	0,8584	0,7378	0,879
113	1000	O	D	10	O	0,8921	0,8941	0,8904	0,8656	0,7359	0,8995
114	5000	O	D	10	O	0,8994	0,9005	0,8992	0,8664	0,8268	0,9051
115	250	Y	D	10	O	0,888	0,8852	0,8641	0,8446	0,7886	0,8849
116	1000	Y	D	10	O	0,8983	0,8996	0,8855	0,8607	0,8855	0,9001
117	5000	Y	D	10	O	0,9009	0,9014	0,9	0,8617	0,7864	0,9015
118	250	D	O	10	O	0,7103	0,6854	0,7083	0,7582	0,504	0,8022
119	1000	D	O	10	O	0,7296	0,7079	0,7202	0,7917	0,5275	0,8171
120	5000	D	O	10	O	0,7496	0,7165	0,7276	0,8036	0,8071	0,8223
121	250	O	O	10	O	0,6496	0,6482	0,641	0,6724	0,5811	0,6964
122	1000	O	O	10	O	0,6681	0,6643	0,6701	0,6974	0,5841	0,7097
123	5000	O	O	10	O	0,6854	0,6792	0,6849	0,7079	0,714	0,7168
124	250	Y	O	10	O	0,6183	0,628	0,5944	0,6128	0,6404	0,6508
125	1000	Y	O	10	O	0,6411	0,6398	0,6294	0,6404	0,5366	0,6574
126	5000	Y	O	10	O	0,6517	0,6517	0,6514	0,6468	0,6625	0,6634
127	250	D	Y	10	O	0,7714	0,7501	0,77	0,7999	0,2908	0,8388
128	1000	D	Y	10	O	0,79	0,7756	0,7862	0,8322	0,273	0,8562
129	5000	D	Y	10	O	0,8061	0,7867	0,7937	0,8442	0,8405	0,8617
130	250	O	Y	10	O	0,7419	0,7362	0,7337	0,739	0,5816	0,7685
131	1000	O	Y	10	O	0,7586	0,7571	0,7594	0,7587	0,6239	0,7862
132	5000	O	Y	10	O	0,772	0,7688	0,7716	0,7658	0,7532	0,7918
133	250	Y	Y	10	O	0,7407	0,7339	0,7092	0,7027	0,684	0,7462
134	1000	Y	Y	10	O	0,7545	0,7542	0,7462	0,7263	0,6933	0,7642
135	5000	Y	Y	10	O	0,7603	0,7608	0,7597	0,7305	0,7052	0,7655
136	250	D	D	25	O	0,8807	0,875	0,8817	0,901	0,8068	0,9189
137	1000	D	D	25	O	0,8917	0,8898	0,8923	0,9228	0,8373	0,939
138	5000	D	D	25	O	0,9022	0,8994	0,8995	0,9209	0,9018	0,9486
139	250	O	D	25	O	0,8644	0,8719	0,8714	0,8578	0,7477	0,8613
140	1000	O	D	25	O	0,88	0,8891	0,8883	0,8425	0,7492	0,8893
141	5000	O	D	25	O	0,8961	0,9002	0,8991	0,819	0,7896	0,9061

Tablo 4.4 – Devamı

SNR	N	BSIZ_K	PRE	DS	BLIK	C5.0	CHAID	C&RT	TAN	MB	DVM
142	250	Y	D	25	O	0,8715	0,8784	0,8587	0,8236	0,8162	0,8571
143	1000	Y	D	25	O	0,8922	0,8971	0,8825	0,8113	0,8932	0,8817
144	5000	Y	D	25	O	0,9	0,901	0,8998	0,792	0,7906	0,8989
145	250	D	O	25	O	0,7227	0,6995	0,7145	0,8195	0,5074	0,842
146	1000	D	O	25	O	0,7445	0,7185	0,724	0,8487	0,4995	0,8707
147	5000	D	O	25	O	0,7631	0,7167	0,7293	0,8651	0,8692	0,8805
148	250	O	O	25	O	0,6415	0,6432	0,6458	0,6963	0,511	0,6886
149	1000	O	O	25	O	0,6633	0,6654	0,675	0,7148	0,5016	0,7138
150	5000	O	O	25	O	0,6823	0,6792	0,6831	0,7187	0,7243	0,7228
151	250	Y	O	25	O	0,6125	0,6228	0,5906	0,6156	0,6339	0,6434
152	1000	Y	O	25	O	0,6361	0,6362	0,627	0,6443	0,5017	0,6591
153	5000	Y	O	25	O	0,6493	0,6488	0,6491	0,6437	0,6648	0,6632
154	250	D	Y	25	O	0,7766	0,7554	0,7757	0,8488	0,306	0,8695
155	1000	D	Y	25	O	0,7978	0,7809	0,787	0,8729	0,2502	0,8965
156	5000	D	Y	25	O	0,8136	0,7865	0,794	0,8843	0,7657	0,9051
157	250	O	Y	25	O	0,7284	0,7225	0,7304	0,7469	0,3764	0,7499
158	1000	O	Y	25	O	0,7483	0,7522	0,7593	0,7548	0,3584	0,7841
159	5000	O	Y	25	O	0,7686	0,7687	0,7716	0,7517	0,7454	0,7944
160	250	Y	Y	25	O	0,7317	0,7243	0,7064	0,6865	0,6794	0,7396
161	1000	Y	Y	25	O	0,7494	0,7477	0,742	0,6994	0,7165	0,761
162	5000	Y	Y	25	O	0,7598	0,7593	0,7588	0,6944	0,6895	0,7651
169	250	Y	D	5	Y	0,9116	0,9117	0,9096	0,9084	0,8912	0,9247
170	1000	Y	D	5	Y	0,9223	0,9247	0,9231	0,9183	0,8912	0,9326
171	5000	Y	D	5	Y	0,9274	0,9264	0,9265	0,915	0,8904	0,9343
178	250	Y	O	5	Y	0,7933	0,7929	0,7871	0,8071	0,8286	0,8341
179	1000	Y	O	5	Y	0,8113	0,8096	0,812	0,8159	0,8262	0,8354
180	5000	Y	O	5	Y	0,8221	0,8172	0,8214	0,8133	0,8217	0,8373
187	250	Y	Y	5	Y	0,8377	0,8368	0,833	0,8422	0,8502	0,8638
188	1000	Y	Y	5	Y	0,8521	0,852	0,852	0,8567	0,852	0,8715
189	5000	Y	Y	5	Y	0,861	0,8581	0,8605	0,8585	0,855	0,8731
196	250	Y	D	10	Y	0,9123	0,9159	0,9129	0,8997	0,8822	0,9225
197	1000	Y	D	10	Y	0,9215	0,9236	0,9223	0,9048	0,8824	0,9334
198	5000	Y	D	10	Y	0,9279	0,9271	0,9265	0,902	0,8799	0,9366
205	250	Y	O	10	Y	0,7978	0,7957	0,7913	0,8102	0,84	0,8354
206	1000	Y	O	10	Y	0,8133	0,8121	0,8149	0,8244	0,8403	0,8427
207	5000	Y	O	10	Y	0,8237	0,8188	0,8226	0,8209	0,8361	0,8435
214	250	Y	Y	10	Y	0,8364	0,8374	0,8369	0,8416	0,8522	0,865
215	1000	Y	Y	10	Y	0,8517	0,8514	0,8542	0,8561	0,8536	0,8744
216	5000	Y	Y	10	Y	0,8615	0,8585	0,8609	0,8558	0,8566	0,8771
223	250	Y	D	25	Y	0,9104	0,9167	0,9156	0,8951	0,876	0,9182
224	1000	Y	D	25	Y	0,9165	0,9223	0,9218	0,8948	0,8795	0,9303
225	5000	Y	D	25	Y	0,9265	0,9271	0,9264	0,8851	0,8699	0,9364
232	250	Y	O	25	Y	0,7908	0,794	0,7953	0,8103	0,8441	0,8304
233	1000	Y	O	25	Y	0,8049	0,8119	0,8158	0,8317	0,8475	0,8427
234	5000	Y	O	25	Y	0,8215	0,8183	0,8229	0,8286	0,8451	0,8462
241	250	Y	Y	25	Y	0,8347	0,8357	0,8397	0,8369	0,8528	0,8612
242	1000	Y	Y	25	Y	0,8437	0,8504	0,8548	0,8551	0,8541	0,8741
243	5000	Y	Y	25	Y	0,8607	0,8589	0,8617	0,8521	0,8571	0,8793

Tablo 4.5. Yöntemlerin Senaryo Bazlı DUY Kriterine Göre Sonuçları

SNR	N	BSIZ_K	PRE	DS	BLIK	C5.0	CHAID	C&RT	TAN	MB	DVM
1	250	D	D	5	D	0,0124	0,0051	0,1309	0,1322	0,0225	0,0495
2	1000	D	D	5	D	0,0191	0,0225	0,0904	0,1912	0,0066	0,0113
3	5000	D	D	5	D	0,0033	0,0003	0,0046	0,2659	0,1742	0,001
4	250	O	D	5	D	0,0149	0,0073	0,0917	0,1243	0,0203	0,0175
5	1000	O	D	5	D	0,0063	0,0058	0,0529	0,1581	0,006	0,009
6	5000	O	D	5	D	0	0	0,0009	0,2141	0,0162	0
7	250	Y	D	5	D	0,0157	0,0119	0,0811	0,0875	0,0168	0,006
8	1000	Y	D	5	D	0,0038	0,0013	0,0368	0,1245	0,0007	0,0037
9	5000	Y	D	5	D	0	0	0	0,1396	0	0,0015
10	250	D	O	5	D	0,5224	0,5032	0,547	0,5547	0,4312	0,5925
11	1000	D	O	5	D	0,5714	0,574	0,5717	0,5763	0,5219	0,6021
12	5000	D	O	5	D	0,5835	0,5827	0,5857	0,5978	0,6005	0,6075
13	250	O	O	5	D	0,496	0,4441	0,5317	0,5428	0,3984	0,5717
14	1000	O	O	5	D	0,5683	0,5675	0,5591	0,5608	0,3152	0,5814
15	5000	O	O	5	D	0,5697	0,5718	0,5711	0,5784	0,558	0,585
16	250	Y	O	5	D	0,4616	0,3804	0,5287	0,5257	0,359	0,5599
17	1000	Y	O	5	D	0,5536	0,565	0,5483	0,5487	0,2503	0,5673
18	5000	Y	O	5	D	0,5653	0,5656	0,5626	0,562	0,2753	0,5686
19	250	D	Y	5	D	0,753	0,6089	0,7606	0,7627	0,7519	0,7578
20	1000	D	Y	5	D	0,7534	0,7549	0,7602	0,7639	0,7526	0,755
21	5000	D	Y	5	D	0,7517	0,7547	0,7567	0,7616	0,7695	0,7536

Tablo 4.5 – Devami

SNR	N	BSIZ_K	PRE	DS	BLI_K	C5.0	CHAIID	C&RT	TAN	MB	DVM
22	250	O	Y	5	D	0,7507	0,5616	0,7537	0,7536	0,7488	0,7516
23	1000	O	Y	5	D	0,7515	0,7506	0,7551	0,7592	0,7507	0,7523
24	5000	O	Y	5	D	0,7508	0,7516	0,7535	0,7575	0,7629	0,7514
25	250	Y	Y	5	D	0,7505	0,5168	0,7517	0,7502	0,7487	0,7487
26	1000	Y	Y	5	D	0,7503	0,7444	0,7508	0,7549	0,75	0,7505
27	5000	Y	Y	5	D	0,75	0,75	0,7514	0,7529	0,7501	0,7501
28	250	D	D	10	D	0,0785	0,0462	0,1721	0,1495	0,0341	0,1583
29	1000	D	D	10	D	0,0724	0,0352	0,1253	0,2125	0,0128	0,1616
30	5000	D	D	10	D	0,0369	0,0058	0,0208	0,2715	0,0766	0,0783
31	250	O	D	10	D	0,0595	0,0174	0,1185	0,1268	0,0301	0,1351
32	1000	O	D	10	D	0,0283	0,0112	0,0798	0,171	0,0031	0,1265
33	5000	O	D	10	D	0,0026	0,0006	0,0017	0,2182	0,0011	0,0045
34	250	Y	D	10	D	0,0405	0,0099	0,0975	0,1089	0,0203	0,0745
35	1000	Y	D	10	D	0,0221	0,0033	0,0477	0,1316	0,0011	0,0072
36	5000	Y	D	10	D	0	0	0,0007	0,1783	0	0
37	250	D	O	10	D	0,5567	0,5502	0,5585	0,5693	0,4783	0,6006
38	1000	D	O	10	D	0,5806	0,578	0,5788	0,5926	0,5438	0,6174
39	5000	D	O	10	D	0,5885	0,5846	0,5879	0,6118	0,5916	0,6206
40	250	O	O	10	D	0,528	0,5152	0,534	0,5384	0,3993	0,5637
41	1000	O	O	10	D	0,5656	0,566	0,5568	0,5642	0,3163	0,5802
42	5000	O	O	10	D	0,5721	0,5726	0,5726	0,5821	0,4965	0,5885
43	250	Y	O	10	D	0,4822	0,4318	0,5212	0,519	0,3903	0,5548
44	1000	Y	O	10	D	0,5634	0,5628	0,5456	0,5432	0,2407	0,5646
45	5000	Y	O	10	D	0,5674	0,564	0,562	0,559	0,2549	0,57
46	250	D	Y	10	D	0,7657	0,7286	0,7736	0,774	0,7609	0,7777
47	1000	D	Y	10	D	0,7571	0,758	0,7636	0,7755	0,7519	0,767
48	5000	D	Y	10	D	0,7569	0,7587	0,7598	0,7815	0,7796	0,7609
49	250	O	Y	10	D	0,7552	0,6562	0,7602	0,7631	0,7489	0,7598
50	1000	O	Y	10	D	0,7529	0,7532	0,7575	0,7679	0,7502	0,7586
51	5000	O	Y	10	D	0,7515	0,7531	0,7546	0,7713	0,7593	0,7544
52	250	Y	Y	10	D	0,7506	0,5501	0,7532	0,7523	0,7481	0,7543
53	1000	Y	Y	10	D	0,7508	0,7474	0,7509	0,7587	0,7496	0,7526
54	5000	Y	Y	10	D	0,7507	0,7507	0,7522	0,7595	0,7506	0,7513
55	250	D	D	25	D	0,1419	0,1046	0,1642	0,1728	0,068	0,1631
56	1000	D	D	25	D	0,1625	0,135	0,1513	0,2016	0,0209	0,1763
57	5000	D	D	25	D	0,1343	0,0292	0,0445	0,2324	0,0885	0,2121
58	250	O	D	25	D	0,1142	0,0662	0,143	0,1479	0,0339	0,1244
59	1000	O	D	25	D	0,0932	0,0477	0,078	0,1656	0,0019	0,1336
60	5000	O	D	25	D	0,0182	0,0017	0,007	0,1768	0	0,1592
61	250	Y	D	25	D	0,0651	0,0254	0,1196	0,1	0,0037	0,1227
62	1000	Y	D	25	D	0,0411	0,0205	0,0698	0,138	0,0008	0,1286
63	5000	Y	D	25	D	0,0047	0	0,0008	0,1525	0	0,1706
64	250	D	O	25	D	0,5663	0,569	0,56	0,587	0,4818	0,5938
65	1000	D	O	25	D	0,5765	0,58	0,5806	0,609	0,5586	0,623
66	5000	D	O	25	D	0,589	0,5863	0,5895	0,6274	0,4504	0,6312
67	250	O	O	25	D	0,5508	0,5423	0,5339	0,5567	0,3903	0,5585
68	1000	O	O	25	D	0,5724	0,5684	0,5606	0,5757	0,2708	0,5789
69	5000	O	O	25	D	0,5731	0,5704	0,5732	0,587	0,4049	0,5905
70	250	Y	O	25	D	0,4944	0,4967	0,5183	0,5251	0,26	0,5446
71	1000	Y	O	25	D	0,5632	0,5652	0,5427	0,5464	0,2302	0,5609
72	5000	Y	O	25	D	0,5649	0,562	0,5601	0,5584	0,25	0,5691
73	250	D	Y	25	D	0,7667	0,7643	0,7695	0,778	0,736	0,7786
74	1000	D	Y	25	D	0,767	0,766	0,7682	0,7941	0,7558	0,7811
75	5000	D	Y	25	D	0,7652	0,7643	0,7635	0,8108	0,7689	0,7717
76	250	O	Y	25	D	0,7558	0,7205	0,7593	0,7652	0,7495	0,7646
77	1000	O	Y	25	D	0,7555	0,7564	0,7605	0,7797	0,7495	0,7667
78	5000	O	Y	25	D	0,7531	0,7547	0,7556	0,7912	0,7483	0,7614
79	250	Y	Y	25	D	0,7526	0,6401	0,7547	0,7587	0,7473	0,7564
80	1000	Y	Y	25	D	0,7525	0,7513	0,7491	0,7669	0,7503	0,7589
81	5000	Y	Y	25	D	0,7507	0,7508	0,7519	0,7718	0,7496	0,7557
82	250	D	D	5	O	0,3186	0,2073	0,3568	0,3687	0,2439	0,4755
83	1000	D	D	5	O	0,4386	0,399	0,4502	0,5039	0,4076	0,5862
84	5000	D	D	5	O	0,5309	0,5081	0,51	0,5598	0,5243	0,6419
85	250	O	D	5	O	0,181	0,0829	0,2904	0,281	0,1787	0,2343
86	1000	O	D	5	O	0,2249	0,1502	0,3365	0,3862	0,2182	0,2417
87	5000	O	D	5	O	0,3327	0,2867	0,3641	0,4359	0,3513	0,5721
88	250	Y	D	5	O	0,0922	0,022	0,2281	0,2272	0,166	0,0603
89	1000	Y	D	5	O	0,0726	0,014	0,2222	0,3199	0,0431	0,0678
90	5000	Y	D	5	O	0,0419	0,0012	0,1542	0,3871	0,2564	0,0756
91	250	D	O	5	O	0,6864	0,6824	0,688	0,7114	0,6618	0,7526
92	1000	D	O	5	O	0,7147	0,7044	0,7154	0,742	0,7208	0,7655
93	5000	D	O	5	O	0,7326	0,7168	0,725	0,7516	0,7541	0,7707
94	250	O	O	5	O	0,6633	0,6665	0,644	0,6663	0,6552	0,6956
95	1000	O	O	5	O	0,6728	0,6695	0,6701	0,6859	0,684	0,7012

Tablo 4.5 – Devami

SNR	N	BSIZ_K	PRE	DS	BLI_K	C5.0	CHAIID	C&RT	TAN	MB	DVM
96	5000	O	O	5	O	0,6826	0,6797	0,6832	0,694	0,6985	0,7043
97	250	Y	O	5	O	0,6487	0,6495	0,6066	0,6162	0,6445	0,6507
98	1000	Y	O	5	O	0,6498	0,6515	0,6401	0,6461	0,5761	0,6602
99	5000	Y	O	5	O	0,6515	0,6536	0,6525	0,6519	0,6597	0,662
100	250	D	Y	5	O	0,8232	0,8167	0,8268	0,8324	0,8298	0,8445
101	1000	D	Y	5	O	0,8298	0,8276	0,8297	0,8447	0,837	0,8516
102	5000	D	Y	5	O	0,8362	0,8254	0,8287	0,8496	0,8574	0,8523
103	250	O	Y	5	O	0,794	0,7949	0,7987	0,8092	0,8118	0,8025
104	1000	O	Y	5	O	0,8	0,8019	0,8037	0,8188	0,8253	0,8065
105	5000	O	Y	5	O	0,7991	0,8006	0,8014	0,8229	0,8461	0,8043
106	250	Y	Y	5	O	0,7727	0,7774	0,7793	0,7859	0,8141	0,7762
107	1000	Y	Y	5	O	0,7784	0,7843	0,7834	0,7938	0,7618	0,7773
108	5000	Y	Y	5	O	0,7792	0,7803	0,7824	0,7985	0,8322	0,7752
109	250	D	D	10	O	0,3929	0,3146	0,3856	0,4182	0,1482	0,5811
110	1000	D	D	10	O	0,448	0,44	0,4519	0,5308	0,1395	0,6673
111	5000	D	D	10	O	0,5148	0,4964	0,5054	0,5591	0,4382	0,6994
112	250	O	D	10	O	0,2561	0,1439	0,3025	0,2914	0,1312	0,3265
113	1000	O	D	10	O	0,3348	0,2859	0,3781	0,3381	0,1271	0,4779
114	5000	O	D	10	O	0,4493	0,4256	0,4104	0,3588	0,3094	0,6015
115	250	Y	D	10	O	0,1418	0,0552	0,232	0,2117	0,1423	0,2321
116	1000	Y	D	10	O	0,144	0,0255	0,2418	0,2688	0,0331	0,2745
117	5000	Y	D	10	O	0,1096	0,0033	0,1994	0,2867	0,2103	0,3285
118	250	D	O	10	O	0,7099	0,6937	0,7127	0,7634	0,096	0,7999
119	1000	D	O	10	O	0,729	0,7118	0,7212	0,7927	0,6132	0,8156
120	5000	D	O	10	O	0,7503	0,7193	0,7295	0,8042	0,808	0,8224
121	250	O	O	10	O	0,6587	0,6613	0,6456	0,6745	0,5696	0,6953
122	1000	O	O	10	O	0,6707	0,668	0,6731	0,6973	0,5939	0,7087
123	5000	O	O	10	O	0,6882	0,6826	0,6878	0,709	0,7154	0,7179
124	250	Y	O	10	O	0,646	0,6493	0,6011	0,6151	0,6439	0,6533
125	1000	Y	O	10	O	0,6519	0,6483	0,6343	0,6421	0,5325	0,6591
126	5000	Y	O	10	O	0,6549	0,6548	0,6549	0,6473	0,6632	0,665
127	250	D	Y	10	O	0,8417	0,8329	0,8368	0,8594	0,2891	0,8805
128	1000	D	Y	10	O	0,8507	0,8377	0,8386	0,8853	0,847	0,8875
129	5000	D	Y	10	O	0,8565	0,832	0,8358	0,8955	0,8992	0,8893
130	250	O	Y	10	O	0,8048	0,8053	0,8085	0,8268	0,7775	0,8135
131	1000	O	Y	10	O	0,8056	0,8069	0,8097	0,844	0,7991	0,8124
132	5000	O	Y	10	O	0,8072	0,8046	0,8064	0,8515	0,8696	0,8109
133	250	Y	Y	10	O	0,7783	0,7824	0,7834	0,7961	0,8191	0,7815
134	1000	Y	Y	10	O	0,782	0,7869	0,7848	0,8089	0,7565	0,779
135	5000	Y	Y	10	O	0,7796	0,7813	0,7825	0,8136	0,8445	0,7749
136	250	D	D	25	O	0,4132	0,3685	0,3936	0,5195	0,0689	0,6233
137	1000	D	D	25	O	0,4569	0,4437	0,4513	0,6158	0,0506	0,7158
138	5000	D	D	25	O	0,5042	0,4812	0,49	0,5668	0,0078	0,7731
139	250	O	D	25	O	0,2935	0,2451	0,3176	0,3087	0,1201	0,2977
140	1000	O	D	25	O	0,3449	0,3631	0,3777	0,3214	0,1223	0,4129
141	5000	O	D	25	O	0,4293	0,4728	0,4366	0,3014	0,279	0,5718
142	250	Y	D	25	O	0,1941	0,0986	0,2338	0,2087	0,1049	0,2252
143	1000	Y	D	25	O	0,2182	0,0781	0,2592	0,2328	0,0198	0,3112
144	5000	Y	D	25	O	0,2065	0,0155	0,205	0,234	0,1503	0,4465
145	250	D	O	25	O	0,7205	0,7037	0,7164	0,8257	0,0437	0,8349
146	1000	D	O	25	O	0,7442	0,7228	0,7258	0,8487	0	0,8667
147	5000	D	O	25	O	0,7625	0,7194	0,7302	0,8653	0,8694	0,8797
148	250	O	O	25	O	0,6484	0,6569	0,6519	0,7015	0,2128	0,6884
149	1000	O	O	25	O	0,6675	0,6706	0,6778	0,7157	0,1287	0,713
150	5000	O	O	25	O	0,6837	0,6813	0,6842	0,7191	0,7246	0,723
151	250	Y	O	25	O	0,6368	0,6413	0,5963	0,619	0,6389	0,6454
152	1000	Y	O	25	O	0,6462	0,6442	0,6304	0,644	0,5055	0,6582
153	5000	Y	O	25	O	0,6531	0,6514	0,6516	0,6442	0,6654	0,6637
154	250	D	Y	25	O	0,85	0,8449	0,8417	0,8917	0,1308	0,9159
155	1000	D	Y	25	O	0,8623	0,8486	0,8423	0,9258	0	0,9266
156	5000	D	Y	25	O	0,8689	0,8345	0,8391	0,9448	0,9061	0,9291
157	250	O	Y	25	O	0,8071	0,8088	0,8079	0,8418	0,4222	0,8175
158	1000	O	Y	25	O	0,8131	0,8133	0,8152	0,8678	0,463	0,8176
159	5000	O	Y	25	O	0,8132	0,8108	0,809	0,8763	0,8873	0,8117
160	250	Y	Y	25	O	0,7828	0,7869	0,7861	0,8057	0,8144	0,7876
161	1000	Y	Y	25	O	0,7846	0,7889	0,7874	0,8256	0,7509	0,7785
162	5000	Y	Y	25	O	0,7817	0,7832	0,7844	0,831	0,8543	0,7746
169	250	Y	D	5	Y	0,5718	0,5197	0,5612	0,5516	0,4838	0,6621
170	1000	Y	D	5	Y	0,6603	0,6826	0,6562	0,5914	0,4782	0,716
171	5000	Y	D	5	Y	0,6917	0,7013	0,6862	0,5651	0,4716	0,7309
178	250	Y	O	5	Y	0,7958	0,7994	0,7887	0,8078	0,8289	0,8326
179	1000	Y	O	5	Y	0,8143	0,8134	0,8144	0,8182	0,8287	0,8344
180	5000	Y	O	5	Y	0,8237	0,8186	0,823	0,819	0,8278	0,8377
187	250	Y	Y	5	Y	0,8839	0,8847	0,8829	0,8967	0,9319	0,8945

Tablo 4.5 – Devamı

SNR	N	BSIZ_K	PRE	DS	BLI_K	C5.0	CHAID	C&RT	TAN	MB	DVM
188	1000	Y	Y	5	Y	0,8866	0,8897	0,8902	0,9057	0,9314	0,8992
189	5000	Y	Y	5	Y	0,8917	0,892	0,8923	0,9052	0,9264	0,8989
196	250	Y	D	10	Y	0,5648	0,5834	0,5714	0,5048	0,4548	0,632
197	1000	Y	D	10	Y	0,6338	0,666	0,6428	0,5154	0,4535	0,7103
198	5000	Y	D	10	Y	0,6825	0,6935	0,6795	0,5018	0,4454	0,7396
205	250	Y	O	10	Y	0,8008	0,8047	0,7951	0,8115	0,8433	0,8334
206	1000	Y	O	10	Y	0,8162	0,8173	0,8171	0,8252	0,8416	0,8427
207	5000	Y	O	10	Y	0,8242	0,8209	0,8239	0,8219	0,8372	0,8436
214	250	Y	Y	10	Y	0,8863	0,8889	0,8886	0,9105	0,9436	0,8998
215	1000	Y	Y	10	Y	0,892	0,8945	0,8942	0,9215	0,944	0,9038
216	5000	Y	Y	10	Y	0,8942	0,8941	0,8936	0,9197	0,9391	0,9014
223	250	Y	D	25	Y	0,5592	0,6046	0,5915	0,4965	0,4457	0,6072
224	1000	Y	D	25	Y	0,596	0,6529	0,6403	0,4863	0,4495	0,6857
225	5000	Y	D	25	Y	0,6629	0,6798	0,6714	0,4549	0,4254	0,7374
232	250	Y	O	25	Y	0,7918	0,8027	0,7976	0,8124	0,8472	0,8267
233	1000	Y	O	25	Y	0,8058	0,8169	0,819	0,8319	0,8479	0,8409
234	5000	Y	O	25	Y	0,8222	0,821	0,8245	0,8292	0,8456	0,8462
241	250	Y	Y	25	Y	0,8897	0,8919	0,891	0,9175	0,9506	0,9022
242	1000	Y	Y	25	Y	0,8922	0,8951	0,8946	0,9339	0,9509	0,9039
243	5000	Y	Y	25	Y	0,8972	0,896	0,8953	0,9322	0,9484	0,9015

Tablo 4.6. Yöntemlerin Senaryo Bazlı SEC Kriterine Göre Sonuçları

SNR	N	BSIZ_K	PRE	DS	BLI_K	C5.0	CHAID	C&RT	TAN	MB	DVM
1	250	D	D	5	D	0,8988	0,5521	0,7988	0,9009	0,8992	0,8992
2	1000	D	D	5	D	0,8816	0,8676	0,8315	0,9013	0,8993	0,8992
3	5000	D	D	5	D	0,9017	0,9017	0,8999	0,9026	0,9029	0,9017
4	250	O	D	5	D	0,9025	0,5391	0,7815	0,9041	0,9015	0,9023
5	1000	O	D	5	D	0,9017	0,8856	0,8051	0,9028	0,9017	0,9016
6	5000	O	D	5	D	0,9015	0,9015	0,8961	0,9021	0,9017	0,9015
7	250	Y	D	5	D	0,8993	0,454	0,7565	0,8993	0,8985	0,8991
8	1000	Y	D	5	D	0,9021	0,8465	0,7943	0,9028	0,902	0,902
9	5000	Y	D	5	D	0,9015	0,9015	0,8925	0,9017	0,9015	0,9015
10	250	D	O	5	D	0,5258	0,5083	0,5508	0,5576	0,4205	0,5993
11	1000	D	O	5	D	0,5717	0,5726	0,5736	0,5777	0,5116	0,6044
12	5000	D	O	5	D	0,5824	0,5819	0,5845	0,5973	0,6007	0,6068
13	250	O	O	5	D	0,4945	0,4522	0,5262	0,5385	0,3741	0,5651
14	1000	O	O	5	D	0,5659	0,5658	0,5583	0,5604	0,3159	0,5818
15	5000	O	O	5	D	0,5703	0,5714	0,5716	0,5775	0,558	0,5844
16	250	Y	O	5	D	0,4617	0,3757	0,5229	0,5239	0,3639	0,5576
17	1000	Y	O	5	D	0,5563	0,5619	0,5465	0,5477	0,259	0,5668
18	5000	Y	O	5	D	0,565	0,5634	0,5622	0,5616	0,2771	0,5687
19	250	D	Y	5	D	0,1381	0,1276	0,3148	0,3222	0,1316	0,328
20	1000	D	Y	5	D	0,1299	0,1792	0,3588	0,3969	0,1597	0,3524
21	5000	D	Y	5	D	0,1275	0,3056	0,3656	0,4621	0,4488	0,4814
22	250	O	Y	5	D	0,0954	0,105	0,2885	0,2823	0,1157	0,2702
23	1000	O	Y	5	D	0,0522	0,0813	0,3084	0,3609	0,0406	0,2278
24	5000	O	Y	5	D	0,0253	0,0933	0,2446	0,4171	0,2696	0,2758
25	250	Y	Y	5	D	0,0834	0,0971	0,2918	0,2667	0,0929	0,2031
26	1000	Y	Y	5	D	0,0234	0,0313	0,2479	0,3435	0,0072	0,1379
27	5000	Y	Y	5	D	0,0049	0,0048	0,1574	0,3974	0,0141	0,1065
28	250	D	D	10	D	0,9053	0,7596	0,8567	0,907	0,9033	0,9085
29	1000	D	D	10	D	0,9029	0,9024	0,8157	0,9074	0,902	0,9031
30	5000	D	D	10	D	0,9018	0,9017	0,8937	0,9069	0,903	0,9017
31	250	O	D	10	D	0,9035	0,6568	0,8323	0,905	0,9021	0,9054
32	1000	O	D	10	D	0,9028	0,899	0,8078	0,9057	0,9024	0,9031
33	5000	O	D	10	D	0,9014	0,9014	0,8951	0,9046	0,9014	0,9014
34	250	Y	D	10	D	0,9038	0,5086	0,8031	0,9045	0,9023	0,9035
35	1000	Y	D	10	D	0,9019	0,8767	0,7652	0,9028	0,9016	0,9017
36	5000	Y	D	10	D	0,9014	0,9014	0,8951	0,9025	0,9014	0,9014
37	250	D	O	10	D	0,5590	0,5495	0,5551	0,5662	0,4728	0,5981
38	1000	D	O	10	D	0,5826	0,5749	0,5798	0,5908	0,5424	0,6164
39	5000	D	O	10	D	0,5868	0,5831	0,5872	0,6109	0,5838	0,6206
40	250	O	O	10	D	0,5255	0,5118	0,5354	0,5399	0,3838	0,5665
41	1000	O	O	10	D	0,5681	0,5643	0,5562	0,5633	0,3176	0,5806
42	5000	O	O	10	D	0,5722	0,5702	0,5721	0,5819	0,5008	0,5884
43	250	Y	O	10	D	0,4984	0,4211	0,5211	0,5223	0,3866	0,5622
44	1000	Y	O	10	D	0,5663	0,5633	0,5455	0,5429	0,2604	0,5648
45	5000	Y	O	10	D	0,5645	0,5634	0,5608	0,5599	0,2547	0,5703
46	250	D	Y	10	D	0,2483	0,2217	0,3466	0,3367	0,2004	0,3718
47	1000	D	Y	10	D	0,2869	0,302	0,38	0,389	0,193	0,452
48	5000	D	Y	10	D	0,3525	0,4056	0,418	0,4412	0,3902	0,543
49	250	O	Y	10	D	0,1773	0,1522	0,3156	0,3128	0,126	0,3211

Tablo 4.6 – Devami

SNR	N	BSIZ_K	PRE	DS	BLIK	C5.0	CHAID	C&RT	TAN	MB	DVM
50	1000	O	Y	10	D	0,1328	0,1511	0,3274	0,348	0,0352	0,3951
51	5000	O	Y	10	D	0,0891	0,1872	0,2984	0,3875	0,1358	0,4596
52	250	Y	Y	10	D	0,1294	0,1125	0,2985	0,2782	0,1109	0,3223
53	1000	Y	Y	10	D	0,0687	0,0576	0,2648	0,3211	0,0012	0,3276
54	5000	Y	Y	10	D	0,0137	0,0119	0,1816	0,3595	0,0032	0,2851
55	250	D	D	25	D	0,9065	0,8776	0,9	0,906	0,9026	0,9092
56	1000	D	D	25	D	0,9045	0,903	0,815	0,9102	0,9012	0,9107
57	5000	D	D	25	D	0,9029	0,9022	0,8951	0,9214	0,9023	0,9104
58	250	O	D	25	D	0,9039	0,836	0,8499	0,9035	0,9008	0,9019
59	1000	O	D	25	D	0,9026	0,902	0,7669	0,9085	0,9015	0,9062
60	5000	O	D	25	D	0,9017	0,8999	0,8945	0,9144	0,9036	0,905
61	250	Y	D	25	D	0,9043	0,6577	0,7969	0,9037	0,9025	0,9049
62	1000	Y	D	25	D	0,9004	0,8859	0,7621	0,904	0,9001	0,9026
63	5000	Y	D	25	D	0,9019	0,8983	0,8892	0,9076	0,9017	0,9024
64	250	D	O	25	D	0,5621	0,5637	0,5574	0,5831	0,5085	0,595
65	1000	D	O	25	D	0,5743	0,5748	0,5786	0,6062	0,5459	0,6229
66	5000	D	O	25	D	0,5868	0,584	0,5868	0,6264	0,5224	0,6311
67	250	O	O	25	D	0,544	0,5384	0,5214	0,545	0,4031	0,5508
68	1000	O	O	25	D	0,567	0,561	0,5562	0,5704	0,3062	0,5762
69	5000	O	O	25	D	0,5744	0,57	0,5725	0,5875	0,3703	0,5911
70	250	Y	O	25	D	0,5213	0,4946	0,5225	0,5263	0,2816	0,5500
71	1000	Y	O	25	D	0,5625	0,5573	0,5408	0,5443	0,2718	0,5601
72	5000	Y	O	25	D	0,5655	0,5584	0,5591	0,5577	0,2469	0,5689
73	250	D	Y	25	D	0,31	0,3082	0,3311	0,3691	0,2516	0,3609
74	1000	D	Y	25	D	0,3684	0,3585	0,389	0,3918	0,2629	0,4352
75	5000	D	Y	25	D	0,4192	0,422	0,4297	0,4158	0,2954	0,53
76	250	O	Y	25	D	0,2547	0,2323	0,3055	0,3177	0,1278	0,3246
77	1000	O	Y	25	D	0,2762	0,2689	0,3455	0,3429	0,023	0,3564
78	5000	O	Y	25	D	0,2259	0,3014	0,3299	0,3571	0	0,4186
79	250	Y	Y	25	D	0,1742	0,1564	0,2917	0,2839	0,0211	0,2976
80	1000	Y	Y	25	D	0,1467	0,1455	0,2757	0,3069	0,0025	0,3285
81	5000	Y	Y	25	D	0,0528	0,0611	0,1818	0,3276	0,0006	0,3696
82	250	D	D	5	O	0,9172	0,9086	0,9198	0,9235	0,9092	0,9244
83	1000	D	D	5	O	0,9202	0,916	0,9202	0,9275	0,9124	0,9274
84	5000	D	D	5	O	0,9204	0,9146	0,917	0,9284	0,9353	0,9278
85	250	O	D	5	O	0,9113	0,8984	0,914	0,9179	0,9141	0,9115
86	1000	O	D	5	O	0,9062	0,905	0,9074	0,9166	0,9093	0,904
87	5000	O	D	5	O	0,9066	0,9066	0,9083	0,9191	0,9365	0,9055
88	250	Y	D	5	O	0,9047	0,8794	0,9021	0,9108	0,915	0,9029
89	1000	Y	D	5	O	0,904	0,9032	0,9054	0,9101	0,9031	0,9033
90	5000	Y	D	5	O	0,9015	0,9011	0,9031	0,9074	0,9344	0,9012
91	250	D	O	5	O	0,6952	0,6789	0,6925	0,7124	0,5386	0,7609
92	1000	D	O	5	O	0,7161	0,7009	0,7131	0,7429	0,679	0,7675
93	5000	D	O	5	O	0,7302	0,7119	0,7232	0,7497	0,7526	0,7685
94	250	O	O	5	O	0,6631	0,6579	0,6439	0,6627	0,6494	0,6959
95	1000	O	O	5	O	0,6715	0,6687	0,6692	0,6846	0,6796	0,7006
96	5000	O	O	5	O	0,6848	0,6785	0,6854	0,6945	0,6997	0,7058
97	250	Y	O	5	O	0,6557	0,6488	0,6102	0,6198	0,6482	0,6589
98	1000	Y	O	5	O	0,6525	0,6478	0,6397	0,6452	0,577	0,6604
99	5000	Y	O	5	O	0,6563	0,6554	0,6568	0,6538	0,6611	0,6652
100	250	D	Y	5	O	0,551	0,5162	0,5426	0,5631	0,3202	0,6685
101	1000	D	Y	5	O	0,5931	0,5731	0,5906	0,6354	0,4627	0,6920
102	5000	D	Y	5	O	0,6315	0,6064	0,6246	0,656	0,6436	0,7019
103	250	O	Y	5	O	0,4632	0,4326	0,4604	0,4889	0,4354	0,5812
104	1000	O	Y	5	O	0,5575	0,5433	0,5384	0,5422	0,4907	0,6303
105	5000	O	Y	5	O	0,5924	0,5761	0,5875	0,5619	0,5219	0,6484
106	250	Y	Y	5	O	0,3037	0,3315	0,4104	0,4335	0,4095	0,5019
107	1000	Y	Y	5	O	0,4584	0,4784	0,5073	0,4976	0,3499	0,5942
108	5000	Y	Y	5	O	0,5712	0,5721	0,5627	0,5101	0,4541	0,6222
109	250	D	D	10	O	0,9287	0,9225	0,9263	0,932	0,9123	0,9404
110	1000	D	D	10	O	0,9315	0,9232	0,9231	0,946	0,9061	0,9441
111	5000	D	D	10	O	0,9322	0,918	0,9196	0,9542	0,9371	0,9451
112	250	O	D	10	O	0,9142	0,9079	0,9178	0,9215	0,9074	0,9176
113	1000	O	D	10	O	0,9124	0,9108	0,9134	0,931	0,9068	0,9167
114	5000	O	D	10	O	0,9103	0,9088	0,9098	0,9378	0,9523	0,9113
115	250	Y	D	10	O	0,9082	0,8956	0,9064	0,915	0,9138	0,9097
116	1000	Y	D	10	O	0,9039	0,9022	0,8983	0,9166	0,9014	0,9044
117	5000	Y	D	10	O	0,9026	0,9015	0,9042	0,9203	0,9394	0,9026
118	250	D	O	10	O	0,7156	0,6881	0,7092	0,757	0,5039	0,8079
119	1000	D	O	10	O	0,7326	0,7075	0,7229	0,7916	0,5285	0,8192
120	5000	D	O	10	O	0,7496	0,7151	0,728	0,8031	0,8069	0,8224
121	250	O	O	10	O	0,6634	0,6584	0,6459	0,6741	0,5805	0,7028
122	1000	O	O	10	O	0,674	0,6676	0,673	0,6985	0,5821	0,7123
123	5000	O	O	10	O	0,6855	0,6775	0,685	0,707	0,713	0,7162

Tablo 4.6 – Devami

SNR	N	BSIZ_K	PRE	DS	BLIK	C5.0	CHAID	C&RT	TAN	MB	DVM
124	250	Y	O	10	O	0,6473	0,6448	0,5985	0,6148	0,6414	0,6569
125	1000	Y	O	10	O	0,6493	0,6437	0,6344	0,6414	0,5316	0,6601
126	5000	Y	O	10	O	0,6539	0,6524	0,6526	0,6474	0,6628	0,6638
127	250	D	Y	10	O	0,5537	0,5106	0,5512	0,6133	0,2601	0,7018
128	1000	D	Y	10	O	0,59	0,563	0,5922	0,6695	0,2552	0,7455
129	5000	D	Y	10	O	0,6286	0,5979	0,6182	0,6894	0,6741	0,7607
130	250	O	Y	10	O	0,4816	0,4656	0,4694	0,4809	0,3446	0,5598
131	1000	O	Y	10	O	0,5389	0,5293	0,5332	0,5171	0,3621	0,6292
132	5000	O	Y	10	O	0,5785	0,5679	0,5789	0,53	0,5055	0,6628
133	250	Y	Y	10	O	0,3731	0,3439	0,4011	0,4061	0,3885	0,4898
134	1000	Y	Y	10	O	0,488	0,5017	0,4941	0,4487	0,2522	0,5925
135	5000	Y	Y	10	O	0,5746	0,5692	0,561	0,4606	0,4349	0,6278
136	250	D	D	25	O	0,9303	0,927	0,9249	0,9355	0,9111	0,9537
137	1000	D	D	25	O	0,9356	0,9273	0,9228	0,9597	0,9116	0,9622
138	5000	D	D	25	O	0,9393	0,9208	0,9218	0,9809	0,903	0,9653
139	250	O	D	25	O	0,9183	0,9147	0,918	0,9253	0,9067	0,9207
140	1000	O	D	25	O	0,9169	0,9146	0,915	0,9431	0,9045	0,9203
141	5000	O	D	25	O	0,9146	0,9107	0,912	0,9556	0,9636	0,9165
142	250	Y	D	25	O	0,9082	0,9022	0,9053	0,9155	0,909	0,9108
143	1000	Y	D	25	O	0,9063	0,9038	0,8993	0,9274	0,9011	0,9109
144	5000	Y	D	25	O	0,9035	0,9017	0,9043	0,936	0,9305	0,9075
145	250	D	O	25	O	0,7293	0,7005	0,7174	0,817	0,508	0,8532
146	1000	D	O	25	O	0,7461	0,7166	0,725	0,8493	0,4995	0,8754
147	5000	D	O	25	O	0,7641	0,7152	0,7303	0,8651	0,869	0,8815
148	250	O	O	25	O	0,6487	0,6459	0,6453	0,6946	0,5101	0,6936
149	1000	O	O	25	O	0,665	0,6661	0,6773	0,7145	0,5013	0,7166
150	5000	O	O	25	O	0,6834	0,6786	0,6848	0,7184	0,7241	0,7238
151	250	Y	O	25	O	0,6404	0,6415	0,5928	0,617	0,631	0,6492
152	1000	Y	O	25	O	0,6474	0,6408	0,6324	0,6464	0,5049	0,6625
153	5000	Y	O	25	O	0,6517	0,6497	0,6519	0,6437	0,6647	0,6633
154	250	D	Y	25	O	0,5639	0,5198	0,5662	0,7193	0,266	0,7446
155	1000	D	Y	25	O	0,5997	0,5678	0,5904	0,7308	0,2502	0,8037
156	5000	D	Y	25	O	0,6353	0,5943	0,6139	0,7342	0,5282	0,828
157	250	O	Y	25	O	0,468	0,4549	0,4665	0,5026	0,274	0,507
158	1000	O	Y	25	O	0,5041	0,5129	0,5313	0,5091	0,2616	0,6099
159	5000	O	Y	25	O	0,557	0,5583	0,5739	0,5019	0,493	0,6753
160	250	Y	Y	25	O	0,4333	0,3765	0,4036	0,3959	0,3574	0,4771
161	1000	Y	Y	25	O	0,5054	0,4925	0,4793	0,4194	0,2228	0,5712
162	5000	Y	Y	25	O	0,569	0,5544	0,5521	0,4155	0,42	0,6227
169	250	Y	D	5	Y	0,9436	0,9437	0,9443	0,9524	0,977	0,9442
170	1000	Y	D	5	Y	0,945	0,9443	0,9465	0,9575	0,9766	0,9496
171	5000	Y	D	5	Y	0,946	0,9426	0,9456	0,9615	0,9792	0,9499
178	250	Y	O	5	Y	0,8065	0,8031	0,7922	0,8113	0,8328	0,839
179	1000	Y	O	5	Y	0,8144	0,8112	0,8142	0,8183	0,8278	0,837
180	5000	Y	O	5	Y	0,8222	0,8172	0,8216	0,8173	0,8242	0,8371
187	250	Y	Y	5	Y	0,7098	0,7096	0,682	0,6856	0,6685	0,7598
188	1000	Y	Y	5	Y	0,7419	0,7331	0,7307	0,7139	0,6713	0,7752
189	5000	Y	Y	5	Y	0,7553	0,7436	0,7532	0,7199	0,6837	0,7817
196	250	Y	D	10	Y	0,9482	0,9484	0,9484	0,9598	0,9839	0,9474
197	1000	Y	D	10	Y	0,9492	0,9458	0,9479	0,9696	0,9775	0,9522
198	5000	Y	D	10	Y	0,9483	0,9451	0,9466	0,9724	0,9856	0,9521
205	250	Y	O	10	Y	0,8056	0,8012	0,7929	0,8128	0,8403	0,8403
206	1000	Y	O	10	Y	0,8154	0,8119	0,8165	0,8251	0,8402	0,8436
207	5000	Y	O	10	Y	0,8248	0,818	0,823	0,8215	0,8369	0,8437
214	250	Y	Y	10	Y	0,6964	0,6999	0,6838	0,6687	0,6639	0,7543
215	1000	Y	Y	10	Y	0,7257	0,7195	0,7275	0,6888	0,6618	0,7748
216	5000	Y	Y	10	Y	0,751	0,7396	0,751	0,6908	0,6731	0,7902
223	250	Y	D	25	Y	0,9497	0,9495	0,9492	0,9663	0,987	0,9516
224	1000	Y	D	25	Y	0,9491	0,9471	0,9488	0,9774	0,9672	0,9524
225	5000	Y	D	25	Y	0,9498	0,9472	0,9482	0,9797	0,9887	0,9523
232	250	Y	O	25	Y	0,7966	0,7959	0,7987	0,8116	0,8440	0,8377
233	1000	Y	O	25	Y	0,8074	0,8104	0,8154	0,8324	0,8475	0,8450
234	5000	Y	O	25	Y	0,8223	0,8166	0,8232	0,8284	0,8452	0,8465
241	250	Y	Y	25	Y	0,6774	0,6857	0,69	0,6525	0,6595	0,7371
242	1000	Y	Y	25	Y	0,6979	0,7143	0,7291	0,6747	0,6583	0,7742
243	5000	Y	Y	25	Y	0,7413	0,7366	0,7495	0,6681	0,6655	0,7989

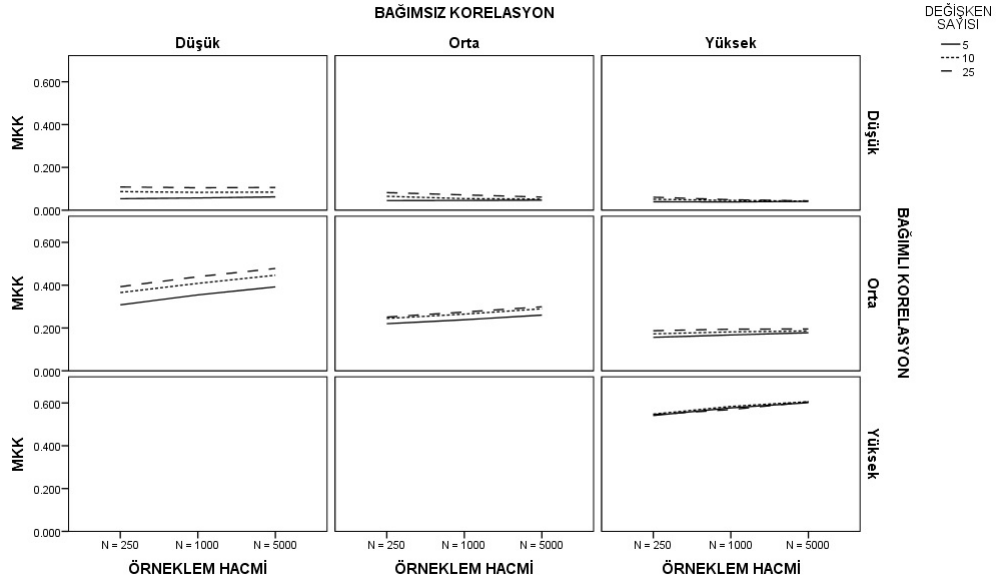
4.2 Grafik Bazlı Sonuçlar

Bu bölümde, karşılaştırma yapılan her bir yöntem aşağıdaki sıra ile grafiksel olarak verilmiş ve grafiklerin altında ilgili yorumlar yapılmıştır.

Karşılaştırılan kombinasyonlara ilişkin kriterlerin grafiksel olarak tanımlanabilmesi ve daha rahat yorumlanabilmesi açısından;

- Her bir yöntem ve her bir model karşılaştırma kriteri için ikişer grafik çizilmiştir.
- Her bir yöntem ve her bir model karşılaştırma kriteri grafiği için ise dört farklı kombinasyon kriterini tanımlayan parametreler kullanılmıştır.

Anlatılmak istenen aşağıda Şekil 4.1 ile kısaca özetlenmiştir;



Şekil 4.1. MKK Kriterine Göre C5.0 Algoritmasının Bağımsız Korelasyon × Bağımlı Korelasyon × Değişken Sayısı × Örneklem Hacmi Senaryo Sonuçları

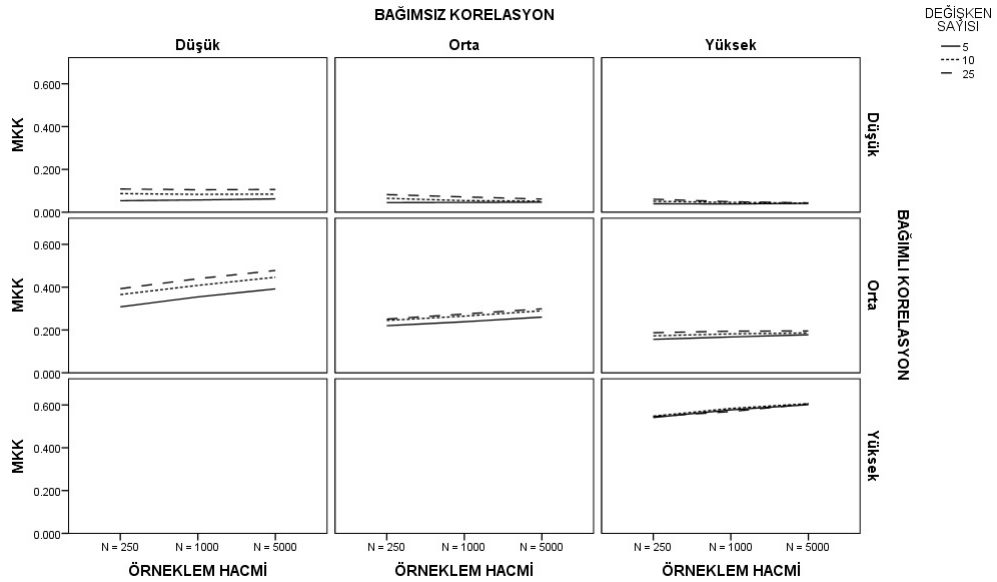
Yukarıda 3×3 'lük bir matrisin her gözesinde değişken sayılarını ifade eden grafikler yer almaktadır. Sadece yukarıdaki şekil ile bir yöntem ve bir model başarı kriterine ilişkin 21 farklı sonuç tanımlanmıştır. Şekil 4.1'e dikkat edildiğinde, (3,1) ve (3,2) hücrelerinde herhangi bir grafiğin olmadığı görülecektir. Bu hücrelere denk gelen 163-168, 172-177, 181-186, 190-195, 199-204, 208-213, 217-222, 226-231 ve 235-240 nolu senaryolara ilişkin veri üretme aşamasında, varyans-kovaryans matrislerinin pozitif tanımlı olmamasından dolayı veriler üretilmemiş ve dolayısıyla da modelleme işlemi gerçekleştirilmemiştir.

Kolonlarda bağımsız değişkenler arasındaki korelasyonlar (BAĞIMSIZ KORELASYON) tanımlanırken, satırlarda bağımsız değişkenlerin bağımlı değişken ile olan korelasyonu (BAĞIMLI KORELASYON) tanımlanmıştır. Her bir hücrenin içerisinde değişken sayıları ve örneklem hacmi ifade edilmiştir.

4.2.1 C5.0 Yöntemine İlişkin Sonuçlar

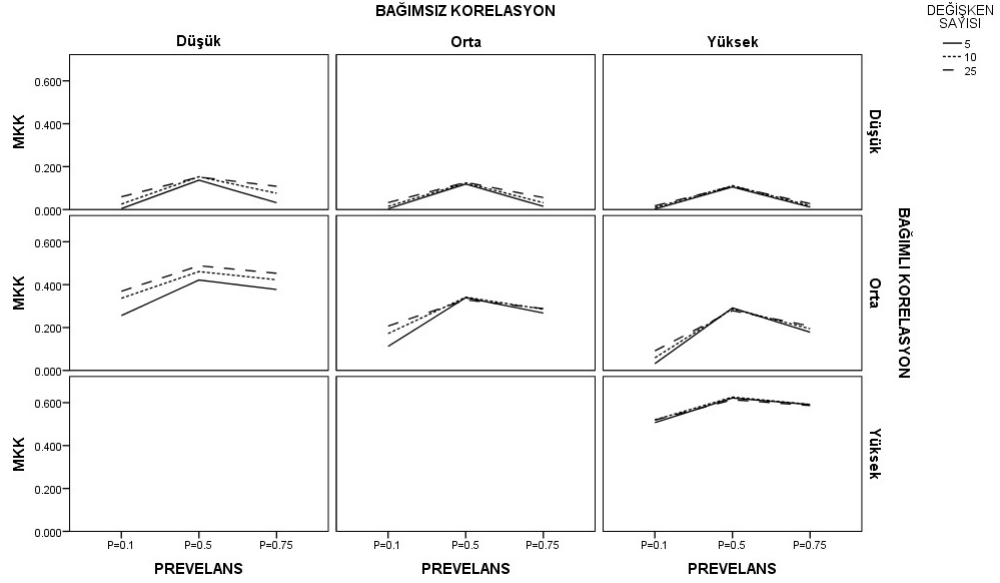
i. MKK Kriterine Göre

a. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi



Şekil 4.2. MKK Kriterine Göre C5.0 Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi Senaryo Sonuçları

b. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans



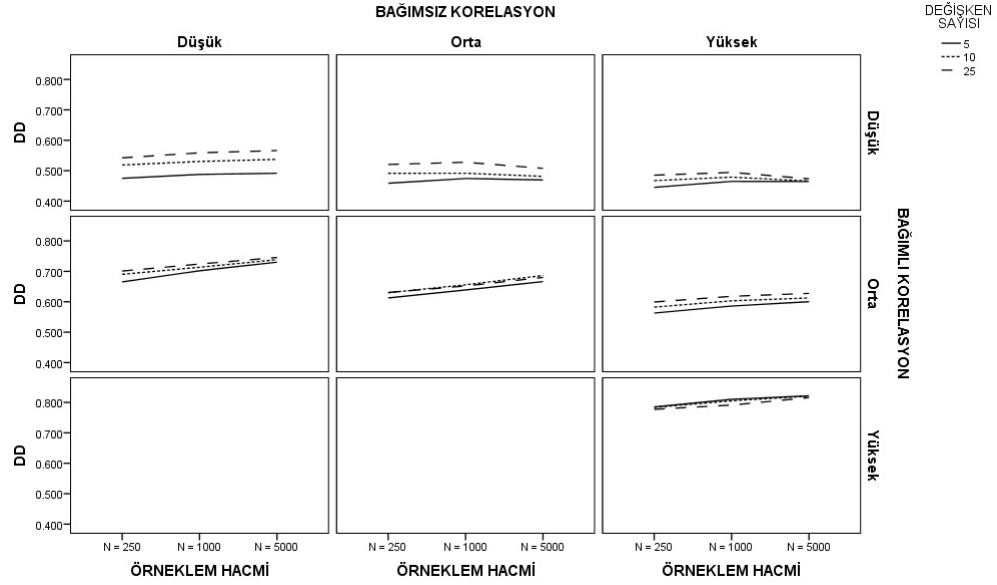
Şekil 4.3. MKK Kriterine Göre C5.0 Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans Senaryo Sonuçları

Yukarıdaki a ve b maddelerine ilişkin Şekil 4.2 ve Şekil 4.3 göz önüne alındığında;

- Örneklem hacminin artması, başarıyı pozitif yönde etkilemektedir,
- Bağımlı değişken kategorilerinin dengeli olması, başarıyı pozitif yönde etkilemiştir,
- $P = 0.1$ prevelans $P = 0.75$ prevelansa göre daha kötü model başarıları göstermektedir,
- Bağımsız değişkenler arasındaki korelasyonun artması, başarıyı negatif yönde etkilemektedir,
- Bağımsız değişken sayısının artması, başarıyı pozitif etkilemektedir,
- Bağımlı değişken ile korelasyonun artması, başarıyı pozitif etkilemektedir,
- Bozulumlardan en az etkilenen kombinasyonun (3,3) hücresindeki yapı olduğu tespit edilmiştir.

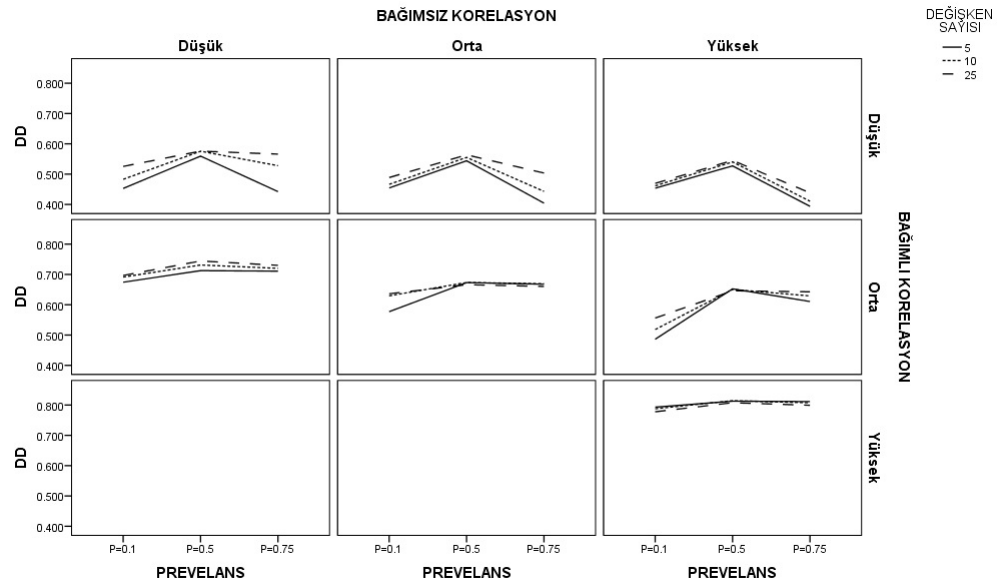
ii. Dengeli Doğruluk Kriterine Göre

a. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi



Şekil 4.4. DD Kriterine Göre C5.0 Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi Senaryo Sonuçları

b. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans



Şekil 4.5. DD Kriterine Göre C5.0 Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans Senaryo Sonuçları

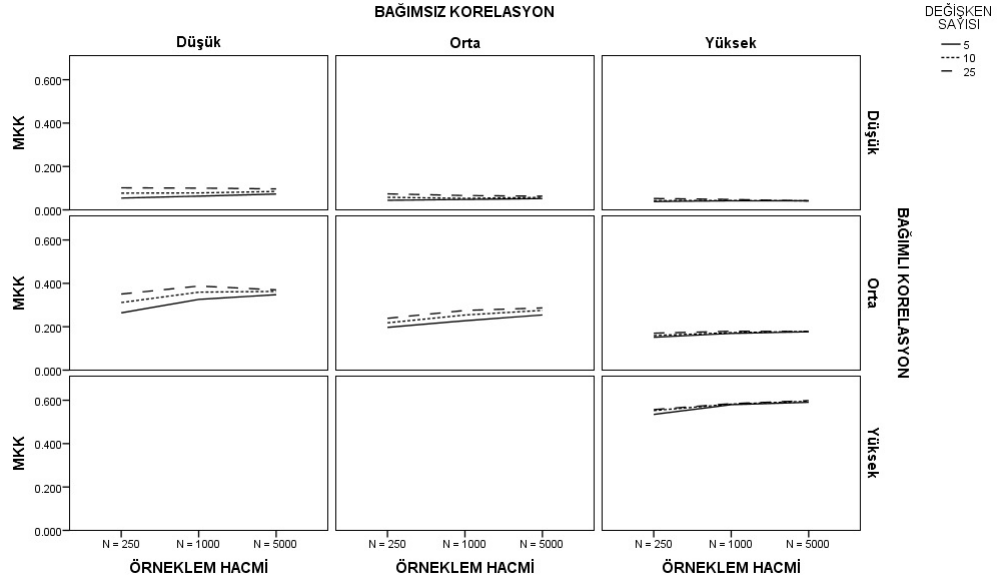
Yukarıdaki a ve b maddelerine ilişkin Şekil 4.4 ve Şekil 4.5 göz önüne alındığında;

- Örneklem hacminin artması, başarıyı pozitif yönde etkilemektedir,
- Bağımlı değişken kategorilerinin dengeli olması, başarıyı pozitif yönde etkilemiştir,
- Bağımsız değişkenler arasındaki korelasyonun artması, başarıyı negatif yönde etkilemektedir,
- Bağımsız değişken sayısının artması, başarıyı pozitif etkilemektedir,
- Bağımlı değişken ile korelasyonun artması, başarıyı pozitif etkilemektedir,
- (2,j) hücreleri incelendiğinde, $P = 0.1$ prevelans sonucunun bağımsız değişkenler arasındaki korelasyon arttıkça başarısının bozulduğu görülmektedir,
- Bozulumlardan en az etkilenen kombinasyonun (3,3) hücresindeki yapı olduğu tespit edilmiştir.

4.2.2 CHAID Yöntemine İlişkin Sonuçlar

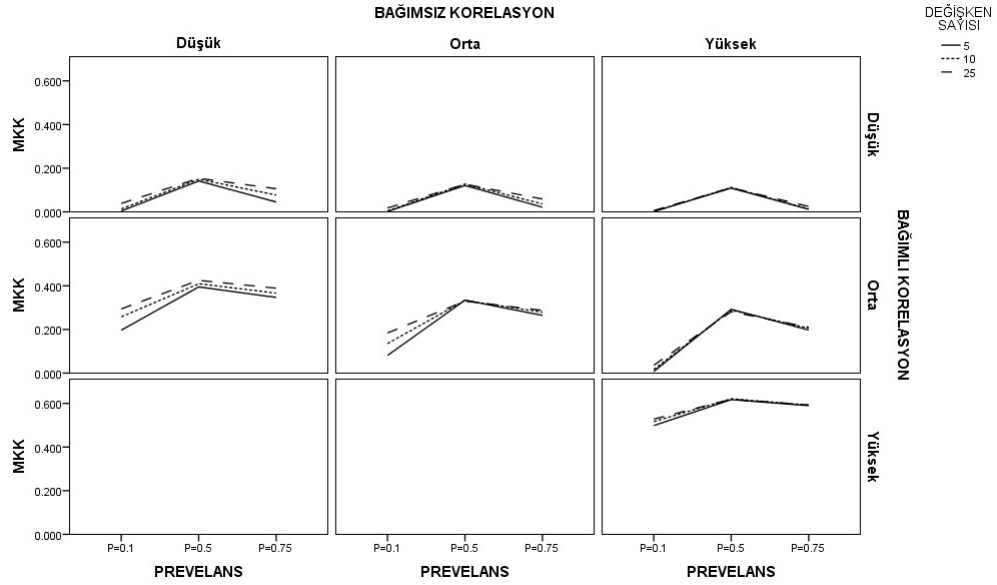
i. MKK Kriterine Göre

a. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi



Şekil 4.6. MKK Kriterine Göre CHAID Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi Senaryo Sonuçları

b. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans



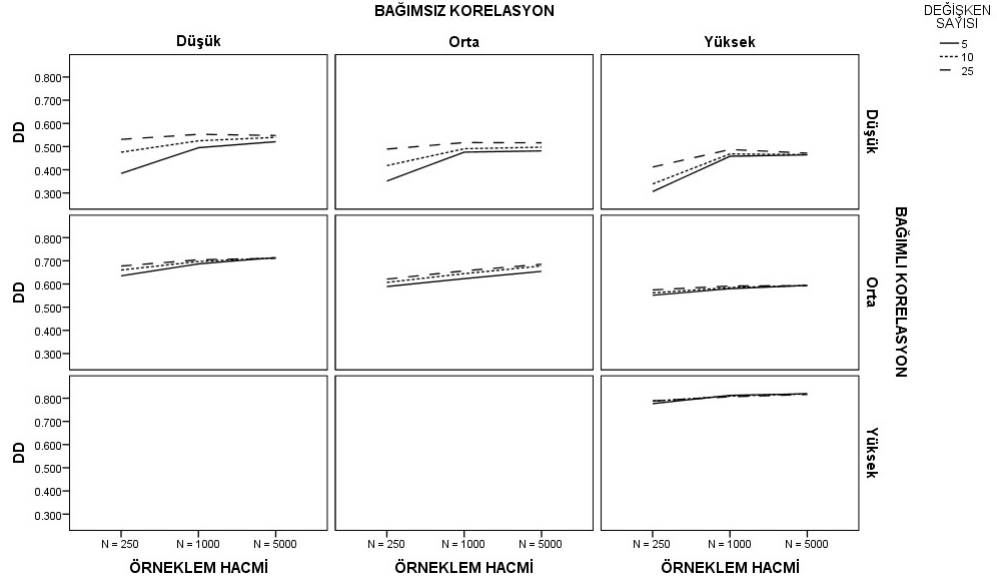
Şekil 4.7. MKK Kriterine Göre CHAID Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans Senaryo Sonuçları

Yukarıdaki a ve b maddelerine ilişkin Şekil 4.6 ve Şekil 4.7 göz önüne alındığında;

- Örneklem hacminin artması (2,j) yapısı hariç, başarıyı etkilememektedir,
- Bağımlı değişken kategorilerinin dengeli olması, başarıyı pozitif yönde etkilemiştir,
- $P = 0.1$ prevelans $P = 0.75$ prevelana göre daha kötü model başarıları göstermektedir,
- Bağımsız değişkenler arasındaki korelasyonun artması (2,j) yapısı hariç, başarıyı etkilememektedir,
- Bağımsız değişken sayısının artması, başarıyı pozitif etkilemektedir,
- Bağımlı değişken ile korelasyonun artması, başarıyı pozitif etkilemektedir,
- Bozulumlardan en az etkilenen kombinasyonun (3,3) hücresindeki yapı olduğu tespit edilmiştir.

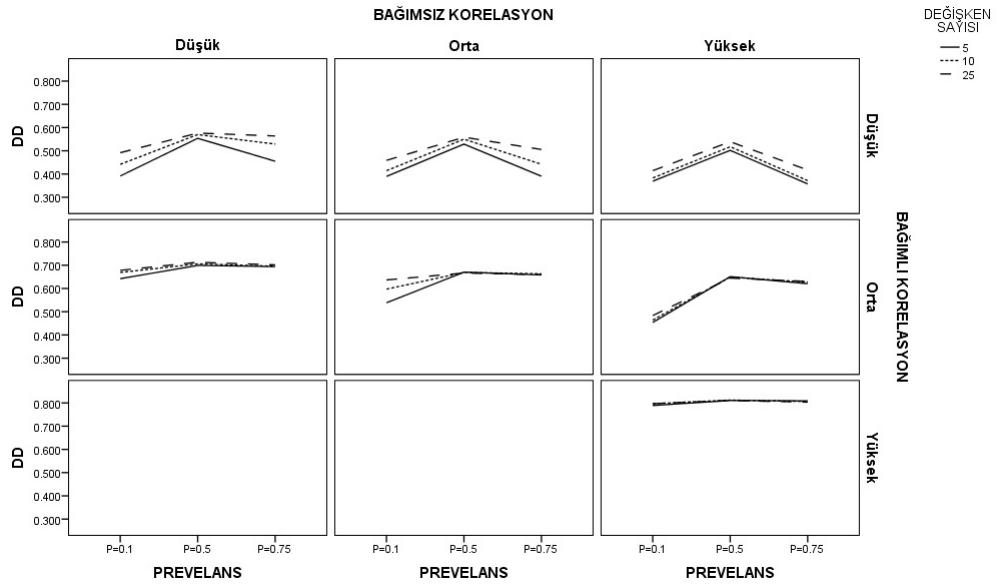
ii. Dengeli Doğruluk Kriterine Göre

a. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi



Şekil 4.8. DD Kriterine Göre CHAID Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi Senaryo Sonuçları

b. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans



Şekil 4.9. DD Kriterine Göre CHAID Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans Senaryo Sonuçları

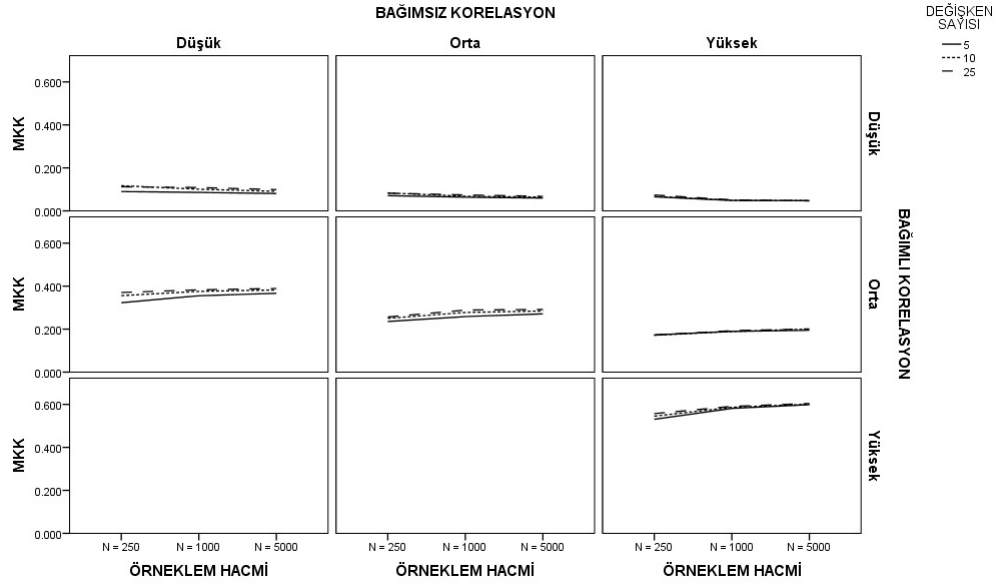
Yukarıdaki a ve b maddelerine ilişkin Şekil 4.8 ve Şekil 4.9 göz önüne alındığında;

- Örneklem hacminin artması, başarıyı pozitif yönde etkilemektedir,
- Bağımlı değişken kategorilerinin dengeli olması, başarıyı pozitif yönde etkilemiştir,
- Bağımsız değişken sayısının artması, başarıyı pozitif etkilemektedir,
- Bağımlı değişken ile korelasyonun artması, başarıyı pozitif etkilemektedir,
- (2,j) hücreleri incelendiğinde, $P = 0.1$ prevelans sonucunun bağımsız değişkenler arasındaki korelasyon arttıkça başarısının bozulduğu görülmektedir,
- Bozulumlardan en az etkilenen kombinasyonun (3,3) hücresindeki yapı olduğu tespit edilmiştir.
- Bağımsız değişkenler ile bağımlı değişken arasındaki korelasyonun artması, bozulumların etkisini azaltmaktadır.

4.2.3 C&RT Yöntemine İlişkin Sonuçlar

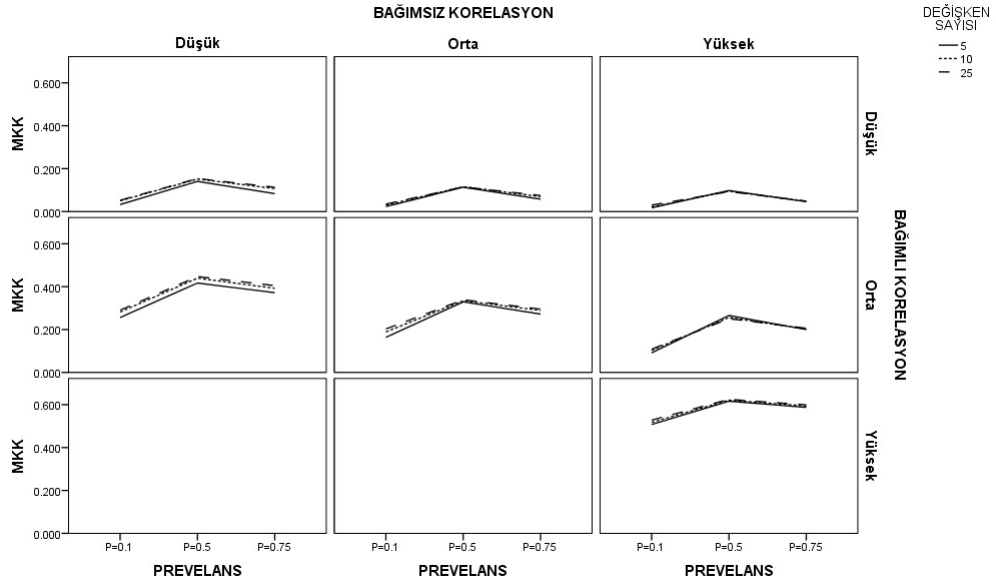
i. MKK Kriterine Göre

a. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi



Şekil 4.10. MKK Kriterine Göre C&RT Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi Senaryo Sonuçları

b. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans



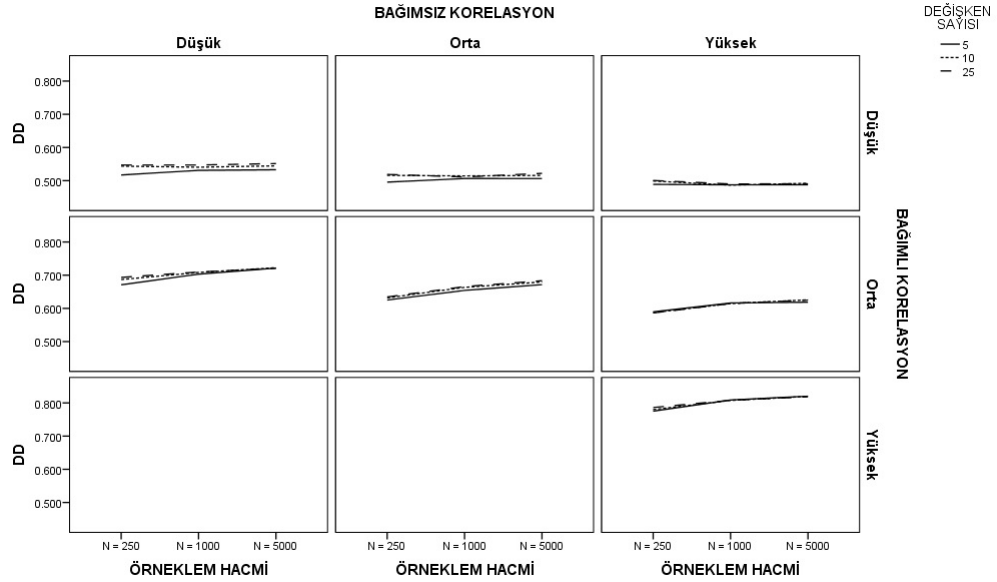
Şekil 4.11. MKK Kriterine Göre C&RT Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans Senaryo Sonuçları

Yukarıdaki a ve b maddelerine ilişkin Şekil 4.10 ve Şekil 4.11 göz önüne alındığında;

- Bağımlı değişken kategorilerinin dengeli olması, başarıyı pozitif yönde etkilemiştir,
- $P = 0.1$ prevelans $P = 0.75$ prevelansa göre daha kötü model başarısı göstermektedir,
- Bağımsız değişkenler arasındaki korelasyonun artması, başarıyı negatif etkilemektedir,
- Bağımlı değişken ile korelasyonun artması, başarıyı pozitif etkilemektedir.

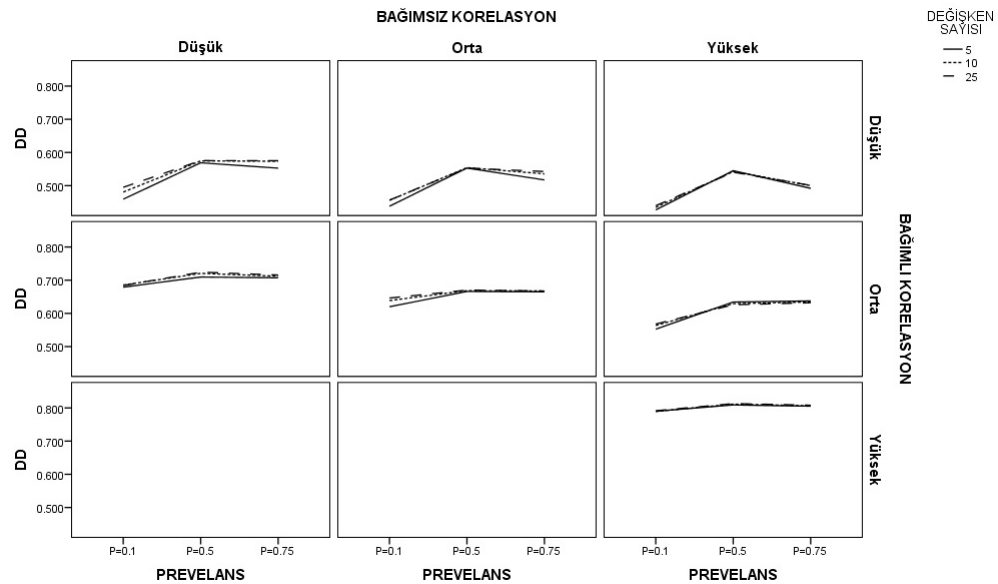
ii. Dengeli Doğruluk Kriterine Göre

a. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi



Şekil 4.12. DD Kriterine Göre C&RT Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi Senaryo Sonuçları

b. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans



Şekil 4.13. DD Kriterine Göre C&RT Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans Senaryo Sonuçları

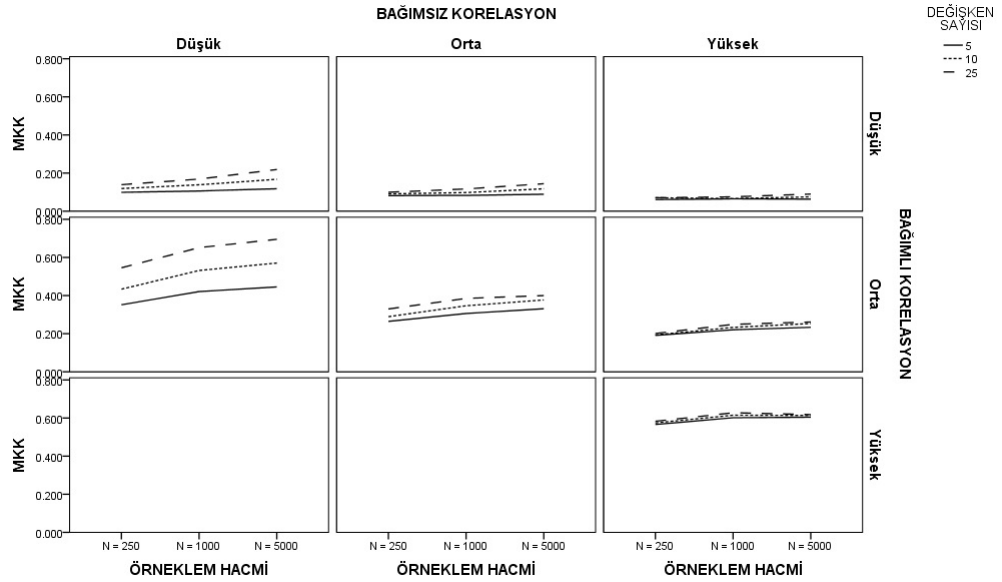
Yukarıdaki a ve b maddelerine ilişkin Şekil 4.12 ve Şekil 4.13 göz önüne alındığında;

- Örneklem hacminin artması, (2,j) yapısı hariç başarıyı etkilememektedir,
- Bağımlı değişken kategorilerinin dengeli olması, başarıyı pozitif yönde etkilemiştir,
- Bağımlı değişken ile korelasyonun artması, başarıyı pozitif etkilemektedir,
- Bağımsız değişkenler arasındaki korelasyonun artması, başarıyı negatif yönde etkilemektedir,
- Bozulumlardan en az etkilenen kombinasyonun (3,3) hücresindeki yapı olduğu tespit edilmiştir.

4.2.4 TAN Yöntemine İlişkin Sonuçlar

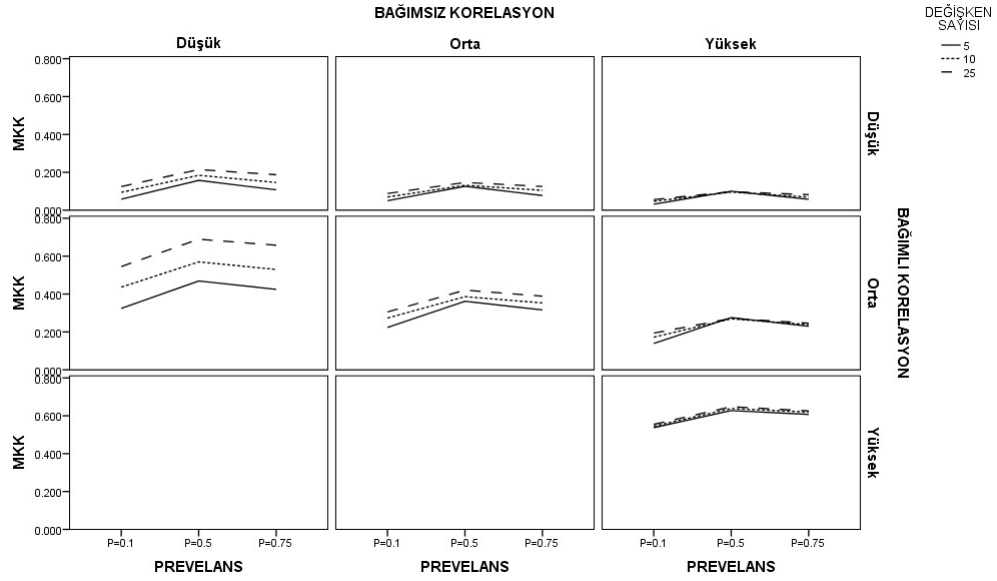
i. MKK Kriterine Göre

a. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi



Şekil 4.14. MKK Kriterine Göre TAN Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi Senaryo Sonuçları

b. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans



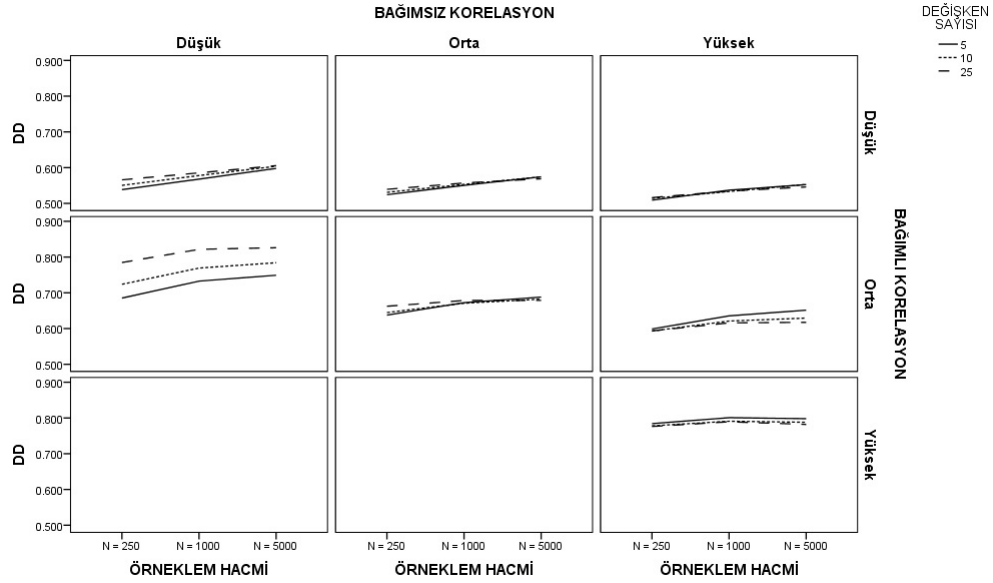
Şekil 4.15. MKK Kriterine Göre TAN Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans Senaryo Sonuçları

Yukarıdaki a ve b maddelerine ilişkin Şekil 4.14 ve Şekil 4.15 göz önüne alındığında;

- Örneklem hacminin artması (2,j) yapısı hariç, başarıyı etkilememektedir,
- Bağımlı değişken kategorilerinin dengeli olması, başarıyı pozitif yönde etkilemiştir,
- $P = 0.1$ prevelans $P = 0.75$ prevelansa göre daha kötü model başarısı göstermektedir,
- Bağımsız değişkenler arasındaki korelasyonun artması, başarıyı negatif yönde etkilemektedir,
- Bağımsız değişken sayısının artması, özellikle (2,1) yapısında başarıyı pozitif etkilemektedir,
- Bağımlı değişken ile korelasyonun artması, başarıyı pozitif etkilemektedir,
- Bozulumlardan en az etkilenen kombinasyonun (3,3) hücresindeki yapı olduğu tespit edilmiştir.

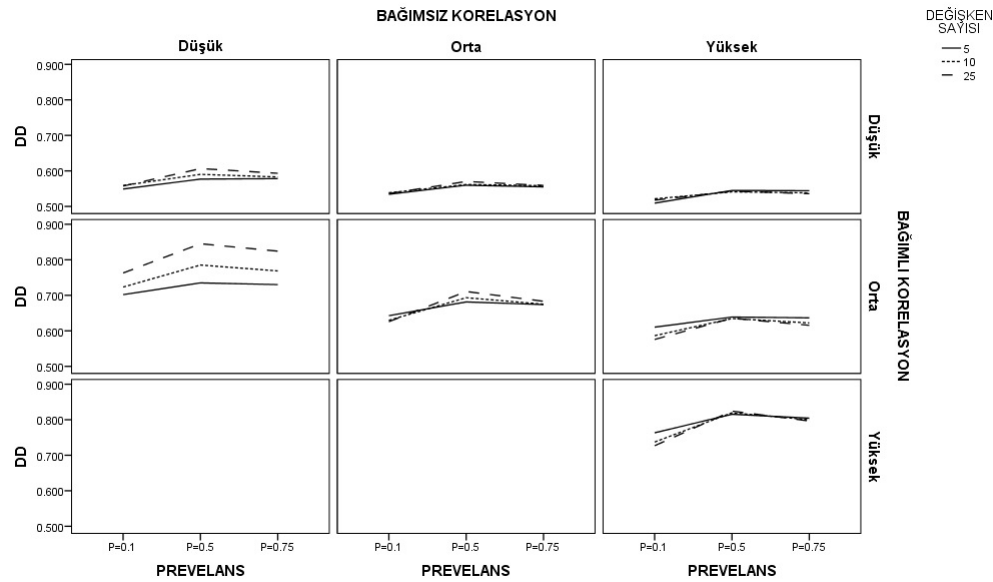
ii. Dengeli Doğruluk Kriterine Göre

a. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi



Şekil 4.16. DD Kriterine Göre TAN Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi Senaryo Sonuçları

b. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans



Şekil 4.17. DD Kriterine Göre TAN Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans Senaryo Sonuçları

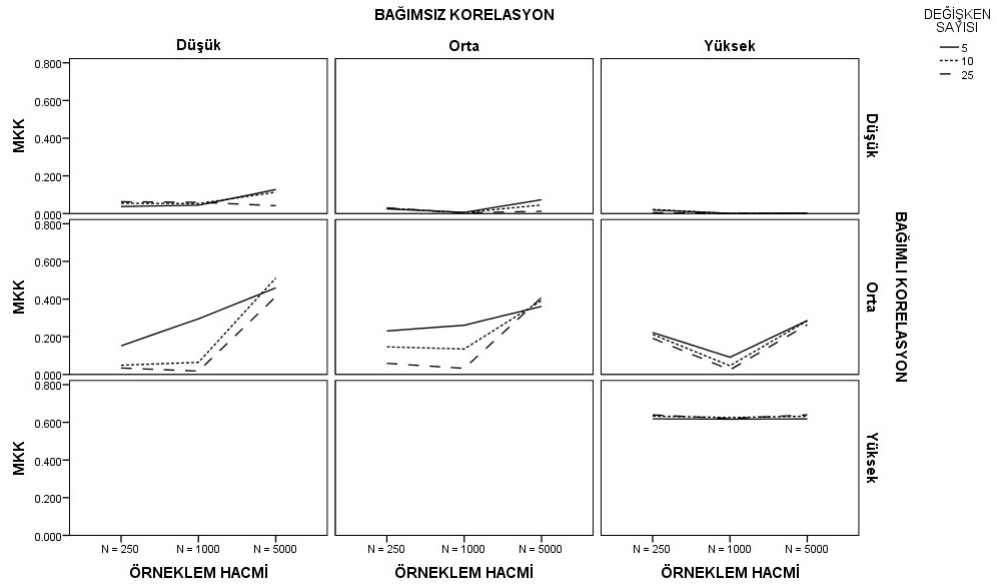
Yukarıdaki a ve b maddelerine ilişkin Şekil 4.16 ve Şekil 4.17 göz önüne alındığında;

- Örneklem hacminin artması, başarıyı pozitif yönde etkilemektedir,
- Bağımlı değişken kategorilerinin dengeli olması, başarıyı pozitif yönde etkilemiştir,
- Bağımsız değişken sayısının artması, (2,1) yapısı hariç etkilememektedir,
- Bağımlı değişken ile korelasyonun artması, başarıyı pozitif etkilemektedir,
- $P = 0.1$ prevelans $P = 0.75$ prevelansa göre daha kötü model başarısı göstermektedir,
- Bozulumlardan en az etkilenen kombinasyonun (3,3) hücresindeki yapı olduğu tespit edilmiştir.

4.2.5 MB Yöntemine İlişkin Sonuçlar

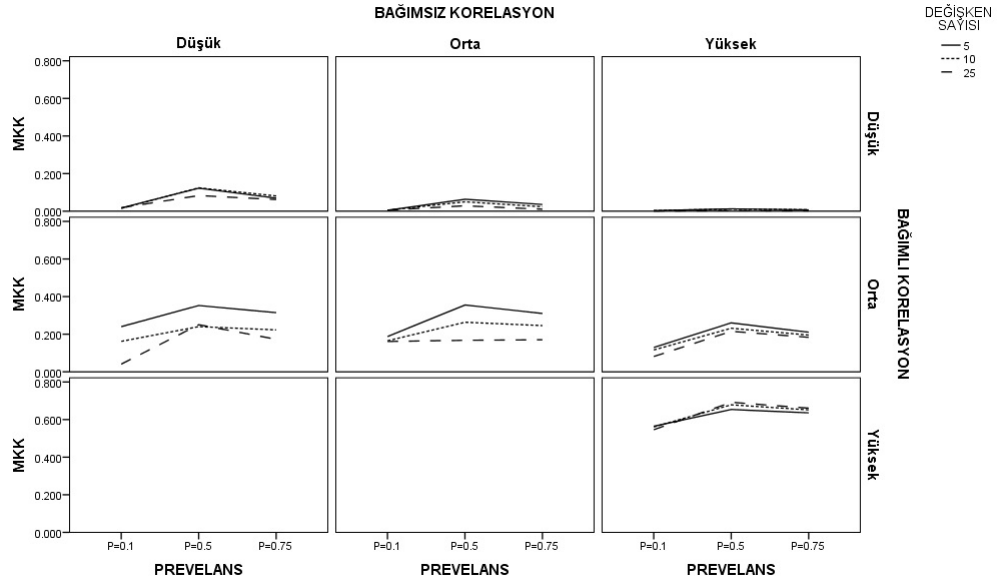
i. MKK Kriterine Göre

a. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi



Şekil 4.18. MKK Kriterine Göre MB Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi Senaryo Sonuçları

b. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans



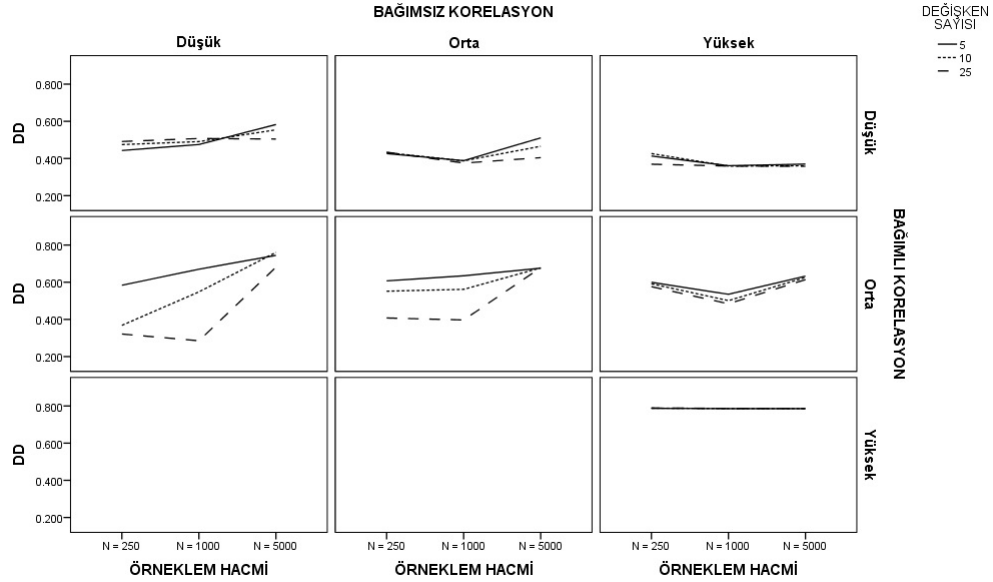
Şekil 4.19. MKK Kriterine Göre MB Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans Senaryo Sonuçları

Yukarıdaki a ve b maddelerine ilişkin Şekil 4.18 ve Şekil 4.19 göz önüne alındığında;

- (1,3) yapısında model başarısı en kötü durumdadır,
- Bağımlı değişken kategorilerinin dengeli olması, başarıyı pozitif yönde etkilemiştir,
- $P = 0.1$ prevelans $P = 0.75$ prevelansa göre daha kötü model başarısı göstermektedir,
- Bağımsız değişkenler arasındaki korelasyonun artması, başarıyı negatif yönde etkilemektedir,
- Bağımsız değişken sayısının azalması, özellikle (2,j) yapısında başarıyı pozitif etkilemektedir,
- Bağımlı değişken ile korelasyonun artması, başarıyı pozitif etkilemektedir,
- Bozulumlardan en az etkilenen kombinasyonun (3,3) hücresindeki yapı olduğu tespit edilmiştir.

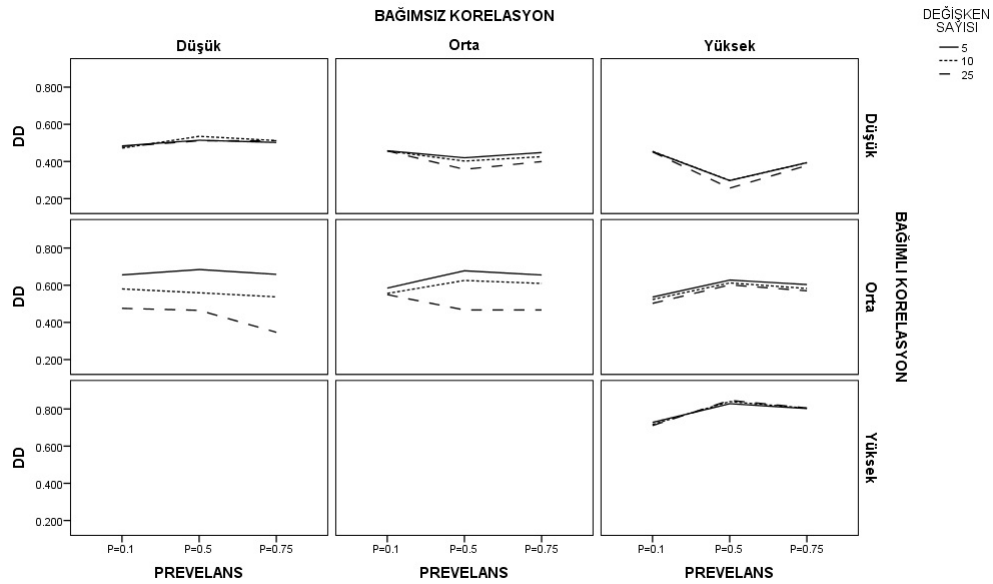
ii. Dengeli Doğruluk Kriterine Göre

a. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi



Şekil 4.20. DD Kriterine Göre MB Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi Senaryo Sonuçları

b. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans



Şekil 4.21. DD Kriterine Göre MB Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans Senaryo Sonuçları

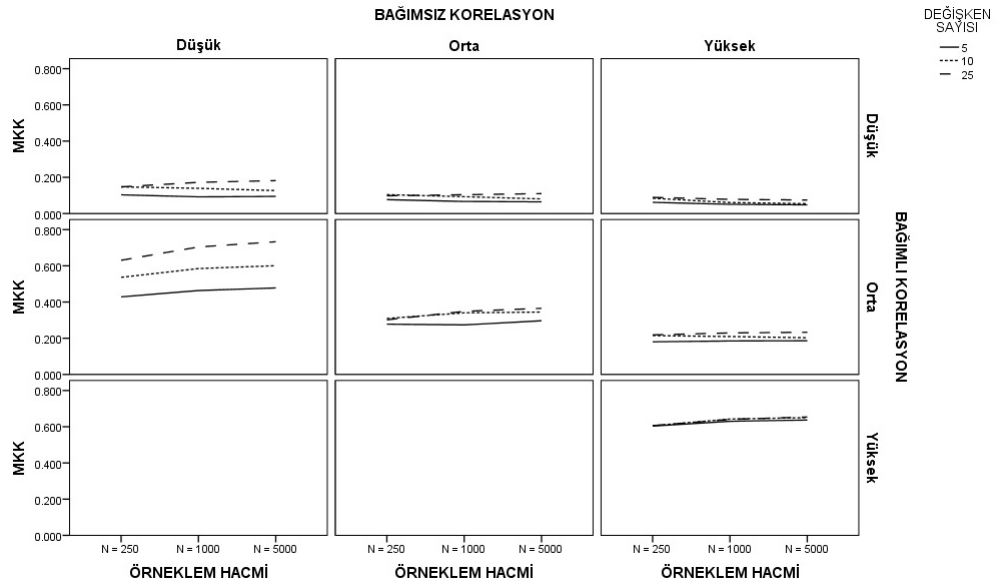
Yukarıdaki a ve b maddelerine ilişkin Şekil 4.20 ve Şekil 4.21 göz önüne alındığında;

- Örneklem hacminin artması, başarıyı pozitif yönde etkilemektedir,
- Bağımlı değişken kategorilerinin dengeli olması, başarıyı pozitif yönde etkilemiştir,
- Bağımsız değişken sayısının azalması, (2,1) ve (2,2) yapıları hariç etkilememektedir,
- Bağımlı değişken ile korelasyonun artması, başarıyı pozitif etkilemektedir,
- Bozulumlardan en az etkilenen kombinasyonun (3,3) hücresindeki yapı olduğu tespit edilmiştir.

4.2.6 DVM Yöntemine İlişkin Sonuçlar

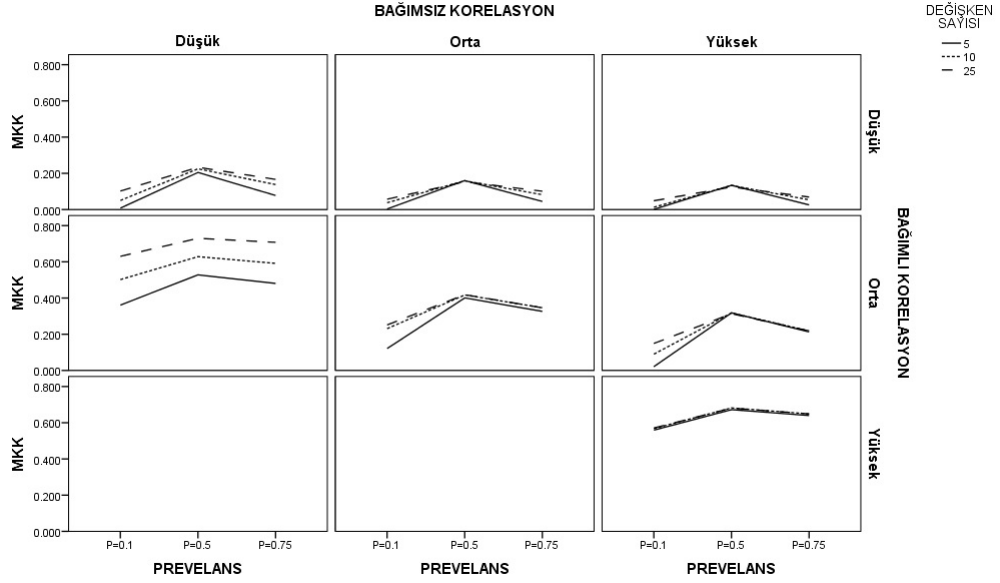
i. MKK Kriterine Göre

a. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi



Şekil 4.22. MKK Kriterine Göre DVM Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi Senaryo Sonuçları

b. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans



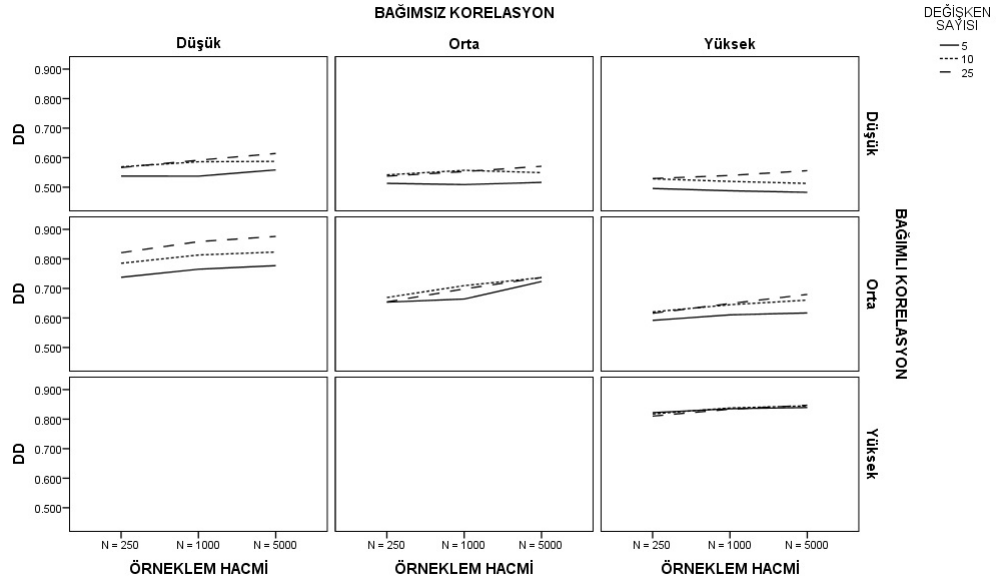
Şekil 4.23. MKK Kriterine Göre DVM Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans Senaryo Sonuçları

Yukarıdaki a ve b maddelerine ilişkin Şekil 4.22 ve Şekil 4.23 göz önüne alındığında;

- Bağımlı değişken kategorilerinin dengeli olması, başarıyı pozitif yönde etkilemiştir,
- $P = 0.1$ prevelans $P = 0.75$ prevelansa göre daha kötü model başarısı göstermektedir,
- Bağımsız değişkenler arasındaki korelasyonun artması, başarıyı negatif yönde etkilemektedir,
- Bağımsız değişken sayısının artması, özellikle (2,1) yapısında başarıyı pozitif etkilemektedir,
- Bağımlı değişken ile korelasyonun artması, başarıyı pozitif etkilemektedir,
- (2,1) yapısı hariç, bozulumlardan çok az etkilendiği tespit edilmiştir.

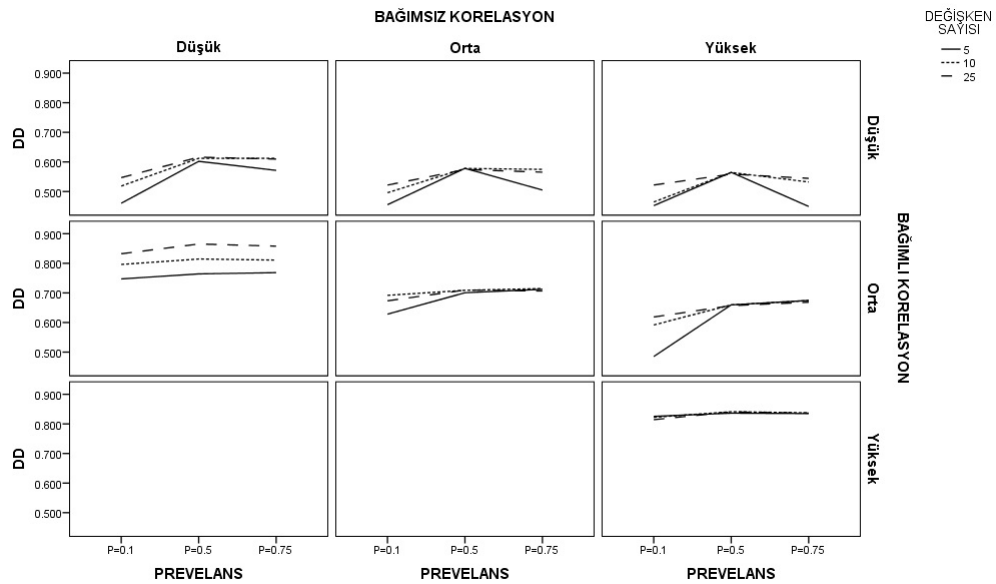
ii. Dengeli Doğruluk Kriterine Göre

a. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi



Şekil 4.24. DD Kriterine Göre DVM Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Örneklem Hacmi Senaryo Sonuçları

b. Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans



Şekil 4.25. DD Kriterine Göre DVM Algoritmasının Bağımsız Korelasyon $\times$ Bağımlı Korelasyon $\times$ Değişken Sayısı $\times$ Prevelans Senaryo Sonuçları

Yukarıdaki a ve b maddelerine ilişkin Şekil 4.24 ve Şekil 4.25 göz önüne alındığında;

- Bağımlı değişken kategorilerinin dengeli olması, başarıyı pozitif yönde etkilemiştir,
- Bağımsız değişken sayısının artması, başarıyı pozitif yönde etkilemektedir,
- Bağımlı değişken ile korelasyonun artması, başarıyı pozitif etkilemektedir,
- $P = 0.1$ prevelans $P = 0.75$ prevelana göre daha kötü model başarısı göstermektedir,
- Bozulumlardan en az etkilenen kombinasyonun (3,3) hücresindeki yapı olduğu tespit edilmiştir,
- Değişken sayısının çok olduğu yapılarda, prevelansın ciddi etkisi ortadan kalkmaktadır.

5 TARTIŞMA

Bölüm 4.1 ve Bölüm 4.2’de yer alan tablo, şekil ve yorumlar genel olarak düşünüldüğünde elde edilen sonuçlar ve bu çalışmayı ileriye götürmesi düşünülen bazı öneriler bu bölümde ele alınmıştır. Buna göre;

- TAN, DVM’ye göre prevelanstan daha az etkilenmektedir,
- MB, örneklem hacminden en fazla etkilenen algoritmadır,
- DVM, TAN’a göre genel olarak daha başarılı sonuçlar vermektedir,
- MKK kriteri dikkate alındığında C&RT, CHAID ve C5.0’dan daha başarılı sonuçlar vermektedir,
- MKK kriteri dikkate alındığında TAN, MB’den daha başarılı sonuçlar vermektedir,
- MKK kriteri dikkate alındığında sıralama, DVM, TAN, C&RT, CHAID, C5.0 ve MB şeklinde olmaktadır,
- DD kriteri dikkate alındığında sıralama, DVM, TAN, CHAID, MB, C&RT ve C5.0 şeklinde olmaktadır,
- Bütün algoritmalar bağımsız değişkenler arasındaki korelasyondan negatif etkilenmektedir.

Tezin hipotezleri ışığında;

- Bayesci ağların küçük örneklemelerde beklenildiği gibi başarılı sonuçlar vermediği,
- Büyük örneklem hacimlerinde karar ağaçları ve DVM’nin daha başarılı sonuçlar verdiği,
- Bağımsız değişkenler arasındaki ilişki miktarı arttıkça bütün yöntemlerin bu durumdan negatif anlamda etkilendiği,
- Bağımsız değişkenler arasındaki ilişki miktarının yüksek olduğu durumlarda TAN ve DVM’nin en iyi sonuçları verdiği,
- Bağımlı değişken prevelansından bütün yöntemlerin başarısının negatif anlamda etkilendiği tespit edilmiştir.

Liratürde daha önce yapılan çalışmalarda;

Huang vd.'nin 2003 yılında yapmış oldukları “Comparing Naïve Bayes, Decision Trees, and DVM with AUC and Accuracy” adlı çalışmada gerçek veri kullanarak; karar ağaçları, DVM ve bayesci ağlar sınıflama yöntemlerini GD ve AUC kriterine göre karşılaştırılmıştır. Bu çalışmada 18 farklı veri kümesi kullanılmıştır. GD başarı kriterine göre çalışmayı değerlendirdiklerinde; DVM, bayesci ağlar ve C4.5 karar ağacı algoritmaları başarılarının çok benzer olduğunu tespit edilmiştir. Yazarlar, DVM sonuçlarının diğer algoritmalara göre çok az daha iyi sonuç gösterdiğini çalışmalarına eklemişlerdir (13).

Amor vd.'nin 2004 yılında yapmış oldukları “Naïve Bayes vs Decision Trees in Intrusion Detection Systems” adlı çalışmada network saldırısı verilerini kullanarak; bayesci ağlar ve karar ağaçları yöntemleri karşılaştırılmıştır. Çalışmada karar ağaçlarının, bayesci ağlara göre daha iyi sonuç verdiği gösterilmiştir (14).

Delen vd.'nin 2005 yılında yapmış oldukları “Predicting breast cancer survivability: a comparison of three data mining methods” adlı çalışmada meme kanseri verisi kullanılarak karar ağaçları, yapay sinir ağları ve lojistik regresyon algoritmaları GD, DUY ve SEC kriterlerine göre karşılaştırılmıştır. C5.0 algoritmasının en iyi sonucu, lojistik regresyonun ise en kötü sonucu verdiğini GD kriterine göre tespit edilmiştir.

Lahiri'nin. 2006 yılında yapmış olduğu “Comparison of Data Mining and Statistical Techniques for Classification Model” adlı yüksek lisans tez çalışmasında 3 farklı veri kümesi lojistik regresyon, karar ağaçları ve yapay sinir ağları algoritmaları kullanılarak karşılaştırılmıştır. Bağımlı değişkenin dağılımının dengesiz olmadı bir veri kümesinde model doğruluklarının değişiklik göstermediği, örneklem hacminin artmasının model başarısına pozitif katkı yaptığı gösterilmiştir (10).

Othman ve Yau'nun 2007 yılında yapmış oldukları “Comparison of different Classification Techniques Using WEKA for Breast Cancer” adlı çalışmada göğüs kanseri verileri kullanılarak; Bayesci Ağlar, Karar Ağaçları, Yapay Sinir Ağları ve k-En Yakın Komşular sınıflama yöntemleri Kappa test istatistiği yardımıyla karşılaştırılmıştır. Bu çalışma sonucunda bayesci ağların en iyi sonucu, k-en yakın komşuların ise en kötü sonucu verdiğini tespit edilmiştir (5).

Amooee vd.'nin 2011 yılında yapmış oldukları “A Comparison Between Data Mining Prediction Algorithms for Fault Detection (Case study: Ahan-pishegan co.)” adlı çalışmada bir fabrikadaki üretim hataları karar ağaçları, bayesci ağlar, lojistik regresyon ve DVM yöntemleri GD ve AUC kriterlerine göre karşılaştırılmıştır. Çalışmada GD kriterine göre en iyiden en kötüye doğru C&RT, DVM, RBF, CHAID ve yapay sinir ağları olduğu tespit edilmiştir.

Tez kapsamında ele alınan senaryoların karşılık geldiği bir çalışmaya literatürde rastlanmamıştır. Çoğu çalışmada gerçek veri kümeleri kullanılarak yapılan sınıflama yöntemlerinde elde edilen sonuçlara göre; karar ağacı algoritmalarının genel olarak daha iyi sonuçlar verdiği, lojistik regresyonunun genel olarak daha kötü sonuçlar verdiği, örneklem hacminin artmasının model başarısına pozitif katkı yaptığı, bağımlı değişkenin dağılımındaki dengesizliğin model başarısına negatif etki yaptığı gösterilmiştir.

Tez kapsamındaki benzetim çalışmasında 6 farklı sınıflama algoritması ve 243 farklı senaryo kullanılmıştır. Bağımsız değişkenlerin tamamı sürekli sayısal ve simetrik yapıdaki çok değişkenli normal dağılıma sahip değişkenlerden oluşmaktadır.

Bu çalışmanın daha da ileriye taşınabilmesi için literatürde yer alan yapay sinir ağları, DVM algoritmasına ait diğer yöntemler, karar listeleri, k-en yakın komşular gibi sınıflama algoritmalarının, simetrik olmayan dağılımdan gelen bağımsız değişkenlerin, kategorik bağımsız değişkenlerin, ilişkisiz bağımsız değişkenlerin (gürültü değişkeni), daha büyük örneklem hacmine sahip yapıların kurgulanması ile hazırlanmış bir benzetim çalışması bu konudaki ileriki çalışmalara örnek olabilir.

KAYNAKLAR

1. Verinin Boyutu. (t.y.). Erişim: 13 Mayıs 2014, <http://searchdatamanagement.techtarget.com/feature/Big-data-growth-increases-data-integration-degree-of-difficulty>
2. Dolgun, M.Ö. (2006). **Büyük Alışveriş Merkezleri İçin Veri Madenciliği Uygulamaları**, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
3. Bichler, M., Kiss, C. (2004). *Comparing Classification Methods for Campaign Management: A Comparison of Logistic Regression, k-Nearest Neighbor, and Decision Tree Induction for Campaign Management*, In: *Proceedings of the Tenth Americas Conference on Information Systems (AMCIS)*, New York.
4. Han, J. and Kamber, M. (2006). *Data Mining: Concepts and Techniques*, The Morgan Kaufmann, Second Edition.
5. Othman, M.F., Yau, T.M.S. (2007). *Comparison of different Classification Techniques Using WEKA for Breast Cancer*, 3rd Kuala Lumpur International Conference on Biomedical Engineering IFMBE Proceedings, Volume 15, Part 13, 520-523.
6. Romero, C., Ventura, S., Hervás, C., Gonzales, P. (2008). *Data mining algorithms to classify students*, I International Conference on Educational Data Mining (EDM), Montreal.
7. Boser, B., Guyon, I., Vapnik, V (1992). *A training algorithm for optimal margin classifiers*, Fifth Annual Workshop on Computational Learning Theory, ACM Press, Pittsburgh.
8. Olson, D.L., Delen, D. (2008). *Advanced Data Mining Techniques*, Springer Publishing.
9. Soria, D., Garibaldi, J.M., Biganzoli, E. (2008). *A Comparison of Three different Methods for Classification of Breast Cancer Data*, Seventh International Conference on Machine Learning and Applications.
10. Lahiri, R. (2006). **Comparison of Data Mining and Statistical Techniques for Classification Model**, Master of Science Thesis, The Department of Information Systems & Decision Sciences, Jadavpur University, India.
11. Reza, E.M., Arash, R., Behrouz M.B. (2009). *Comparison of Classification*

Methods Based on the Type of Attributes and Sample Size, *Journal of Convergence Information Technology*, Volume 4, Number 3.

12. Tjen-Sien, L., Wei-Yin, L., Yu-Shan S. (2000). A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms, *Machine Learning*, 40, 203-229, Kluwer Academic Publishers, Boston.
13. Huang, J., Lu, J., Ling, C.X. (2003). Comparing Naïve Bayes, Decision Trees, and DVM with AUC and Accuracy. *Proceedings of the Third IEEE International Conference on Data Mining*, pp.550.
14. Amor, N.B., Benferhat, S., Elouedi, Z. (2004). Naïve Bayes vs Decision Trees in Intrusion Detection Systems, *ACM Symposium on Applied Computing, Cyprus*.
15. Dolgun, M.Ö., Özdemir, T.G., Oğuz, D. (2009). Veri Madenciliği'nde yapısal olmayan verinin analizi: Metin ve web madenciliği, *İstatistikçiler Dergisi*, 2, 48-58.
16. Larose, D.T. (2006). *Data Mining Methods and Models*, Wiley-Interscience, New Jersey, 322p.
17. Larose, D.T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*, Wiley-Interscience, New Jersey, 222p.
18. Adriaans, P., Zantinge, D. (1996). *Data Mining*, Addison Wesley Longman Limited, Harlow, 159p.
19. Savaş, S., Topaloğlu, N., Yılmaz, M. (2012). Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri, *İstanbul Üniversitesi Fen Bilimleri Dergisi*, 21, 1-23.
20. Ersel, D., Dolgun, M.Ö. (2014). Doğrudan Pazarlama Stratejilerinin Belirlenmesinde Veri Madenciliği Yöntemlerinin Kullanımı, Baskıda, *İstatistikçiler Dergisi: İstatistik&Akrüerya*, Ankara.
21. Tan, P., Steinbach, M., Kumar, V. (2006). *Introduction to Data Mining*, Addison-Wesley, Boston, 769p.
22. Ersel, D. (2012). **Birliktelik Analizinde Özgün Bir Birleşik İlginçlik Ölçümü**, Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
23. Ben-Gal, I. (2007). *Bayesian Networks*, Encyclopedia of Statistics in Quality and Reliability, Ruggeri, F., Faltin, F., Kenett, R. (eds), Wiley & Sons, 1800p.

24. Cheng, J., Bell, D., Liu, W. (1998). *Learning Bayesian Networks from Data: An Efficient Approach Based on Information Theory*, Technical Report, Department of Computer Science, University of Alberta, 41p.
25. Aydoğan, Ü. (2010). ***Destek Vektör Makinalarında Kullanılan Çekirdek Fonksiyonların Sınıflama Performanslarının Karşılaştırılması***, Yüksek Lisans Tezi, Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara.
26. Wang, L. (2005). *Support Vector Machines: Theory and Application*, Studies in Fuzziness and Soft Computing, Volume 177, Springer.
27. He, H. and Garcia, E. A. (2009). *Learning from Imbalanced Data*, *IEEE Transactions on Knowledge and Data Engineering*, 21(9).
28. Menardi, G. and Torelli, N. (2014). Training and assessing classification rules with unbalanced data, *Data Mining and Knowledge Discovery*, 28(1), pp 92-122.
29. Bowes, D., Hall, T. and Gray, D. (2012). *Comparing the performance of fault prediction models which report multiple performance measures: recomputing the confusion matrix*, presented at PROMISE '12, Lund, Sweden.
30. Velez, D.R., White, B.C., Motsinger, A.A., Bush, W.S., Ritchie, M.D., Williams, S.M. and Moore, J.H. (2007). A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction, *Genetic Epidemiology*, vol. 31, no. 4, pp. 306–315.
31. Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M. (2010). *The Balanced Accuracy and Its Posterior Distribution*, 20th International Conference on Pattern Recognition (ICPR), pp. 3121–3124.
32. Pepe, M.S. (2004). *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford Statistical Science Series, 318 pages.
33. Powers, D.M.W. (2007). Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation, *Journal of Machine Learning Technologies*, 2(1): 37–63.
34. Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A.F., Nielsen, H (2000). Assessing the accuracy of prediction algorithms for classification: an overview, *Bioinformatics*, 16, 412–424.
35. Matthews, B.W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme, *Biochimica et Biophysica Acta*, vol. 405, pp.

442-451.

36. Box, G. E. P., Muller, M. E. (1958). A Note on the Generation of Random Normal Deviates, *The Annals of Mathematical Statistics*, 29, no. 2, 610-611. doi:10.1214/aoms/1177706645.
37. Chen, W.H., Hsu, S.H., Shen, H.P. (2005). Application of DVM and ANN for intrusion detection, *Journal of computers and operations research*, Vol. 32, Elsevier, pp. 2617-2634.
38. Sever, H., Raghavan, V.V., Deogun, J.S., and Choubey, S.K. (1997). A Comparison of Classification Methods, *International Conference of Information Sciences – Rough Set 38, Computer Science, Vol. 3*, pp. 371-374.
39. Michie, D., Spiegelhalter, D.J., Taylor, C.C., Campbell, J. (1994). *Machine Learning, Neural and Statistical Classification*, Ellis Horwood.

EK 1. BENZETİM ÇALIŞMASI İÇİN KODLAR

Benzetim çalışması sırasında kullanılan kodlar IBM SPSS Statistics syntax ve script dili ile yazılmış ve kodlar ilgili paket program altında çalıştırılmıştır.

```

DEFINE !kaydet().
!DO !rpt=1 !TO 1000.

matrix.
  compute n = 250.
  compute exact = 0.
  compute r =
{1, 0.20, 0.20, 0.20, 0.20, 0.20 ;
0.20, 1, 0.20, 0.20, 0.20, 0.20 ;
0.20, 0.20, 1, 0.20, 0.20, 0.20 ;
0.20, 0.20, 0.20, 1, 0.20, 0.20 ;
0.20, 0.20, 0.20, 0.20, 1, 0.20 ;
0.20, 0.20, 0.20, 0.20, 0.20, 1}.

  compute rn = nrow(r).
  compute x1 = sqrt(-2*ln(uniform(n,rn)))*cos((2*3.14159265358979)*
uniform(n,rn)).
  compute x1=x1*chol(r).
  compute ones = make(n,1,1).
  compute sigma = (t(x1)*(ident(n)-(1/n)*ones*t(ones))*x1)*(1/(n-1)).
  do if (exact = 1).
    call eigen(r, vc, vl).
    compute sqrtr = vc*sqrt(mdiag(vl))*t(vc).
    call eigen(sigma, vc, vl).
    compute sqrts = vc*sqrt(mdiag(vl))*t(vc).
    compute x1 = x1*inv(sqrts)*sqrtr.
    compute ones = make(n,1,1).
    compute sigma = (t(x1)*(ident(n)-(1/n)*ones*t(ones))*x1)*(1/(n-1)).
  end if.

  *print r/title = "Kitle Matrisi"/format = F16.4.
  *print sigma/title = "Örneklem Matrisi"/format = F16.4.
  *print n/title = "Gözlem Sayısı"/format = F16.0.
save x1 /outfile=* /variables= nr1 to nr6.

```

```
end matrix.
```

```
descriptives variables=nr6
```

```
  /save
```

```
  /statistics=mean stddev variance min max kurtosis skewness.
```

```
compute target=Znr6 > 1.29.
```

```
execute.
```

```
save OUTFILE=!QUOTE(!CONCAT("C:\veriler\senaryo_1\dosya",!rpt,".sav")).
```

```
!DOEND.
```

```
!ENDDEFINE.
```

```
!kaydet.
```

EK 2. MODELLEME ÇALIŞMASI İÇİN KODLAR

Modelleme çalışması sırasında kullanılan kodlar IBM SPSS Modeller syntax dili ile yazılmış ve kodlar ilgili paket program altında çalıştırılmıştır.

```
for i from 1 to 1000
set y = 'C:\veriler\senaryo_1\dosya'><^i ><'.sav'
create statisticsimportnode
set :statisticsimportnode.full_filename = ^y
set :statisticsimportnode.import_names = false
set :statisticsimportnode.import_data = false
connect :statisticsimportnode to :partitionnode
set :partitionnode.set_random_seed = True
set :partitionnode.random_seed = "123456"+1
set 'Type od'.value_mode.nr1 = Read
set 'Type od'.value_mode.nr2 = Read
set 'Type od'.value_mode.nr3 = Read
set 'Type od'.value_mode.nr4 = Read
set 'Type od'.value_mode.nr5 = Read
set Tekrar_No:derivenode.formula_expr = ^i
execute svm_m:svmnode
execute tan_m:bayesnet
execute mb_m:bayesnet
execute crt_m:cartnode
execute c50_m:c50node
execute chaid_m:chaidnode
insert model svm connected between 'Type od' and 'Filter qa'
insert model crt:cartnode connected between svm and 'Filter qa'
insert model cbes:c50node connected between crt and 'Filter qa'
insert model tan connected between cbes and 'Filter qa'
insert model mb connected between tan and 'Filter qa'
insert model chaid:chaidnode connected between mb and 'Filter qa'
execute veri
delete crt
delete chaid
delete cbes
delete svm
delete tan
```

```
delete mb  
delete :statisticsimportnode  
clear outputs  
clear generated palette  
endfor
```