



## Sivas Erzincan Kalkınma Projesi (SEKP) Verilerinin Veri Madencilięi ile Sınıflandırılması ve Kümelenmesi

**Mehmet Ali Alan**

*Yrd. Doç. Dr. Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Fakültesi Yönetim Bilişim Sistemleri Bölümü Sivas, [alan@cumhuriyet.edu.tr](mailto:alan@cumhuriyet.edu.tr)*

### Özet

Bu çalışmada, sınıflandırma ve kümeleme yöntemiyle Sivas Erzincan Kalkınma Projesine (SEKP) ait veriler kullanılarak veri madencilięi yapılmıştır. SEKP verileri ile hem bu verileri en başarılı sınıflandıran algoritma, hem de bu algoritmasının ürettięi sınıflar tespit edilmiştir. Ayrıca bu veriler ile EM algoritmasıyla kümeleme analizi yapılmıştır. Çalışmanın sonucunda BayesNet algoritmasının sınıflandırma oranı en yüksek algoritma olduęu, yedi sınıf ürettięi ortaya çıkmıştır. EM Algoritmasıyla yapılan analizde de verilerin üç ayrı kümede kümelendięi saptanmıştır.

**Anahtar kelimeler:** Veri Madencilięi, Sınıflandırma Analizi, Kümeleme Aalizi, SEKP, Bayesian classifier

### Abstract

In this study, data mining was conducted with data of Sivas Erzincan Development Project (SEDP). The best algorithm classifies most successfully and classes generated by this algorithm were determined with SEDP data. Additionally, clustering analysis was conducted with these data by means of

EM algorithm. As a result of this study, it was emerged that BayesNet is the algorithm with the highest classification rate and generated seven classes. As for the analysis conducted with EM algorithm, it was determined that data were clustered in three clusters.

**Key Words:** Data Mining, Classification Analysis, Clustering Analysis, SEDP, Bayesian classifier

## I. Giriş

Kamu kurumları da özel sektör kurumları gibi veriler tutmaktadır. Hatta bu kurumlar diğer kurumların asla ulaşmayacağı miktarlarda bilgi depolarına sahiptirler. Bu kurumların mevcut bilgileri kâra dönüştürme gibi bir amaca sahip olmadıklarından çoğu zaman bu kaynaklar boşu boşuna tutulmaktadır ya da heba edilmektedirler. Büyük miktarlarda veri tutan bu kamu veritabanları pek çok istatistiksel ve matematiksel bilgiyi ve gizli örüntüyü barındırıyor olabilir. Bu kurumlardan bir tanesi de pek çok çiftçinin bilgisini elinde tutan Tarım Bakanlığı ya da onun taşradaki temsilcilikleridir. Bu çalışmada Sivas Erzincan Kalkınma Projesi (SEKP) kapsamında hibe talebinde bulunan çiftçilere ait veriler kullanılmıştır.

Bu çalışmanın amacı, veri madenciliği tekniğini kullanarak, Sivas Tarım İl Müdürlüğünden elde edilen veriler çerçevesinde analizler yapılmıştır. Bu amaçla adı geçen kurumdan elde edilen veriler, veri madenciliğinin en yaygın kullanılan tekniklerinden “Sınıflandırma” ve “Kümeleme” analizi yöntemi ile analize tabi tutulmuştur.

SEKP, Sivas ve Erzincan illerinin az gelişmiş bölgelerinde, tarımsal verimliliğin ve gelir seviyesinin artırılması ve kırsal göçün önlenmesi ana amacıyla sosyal ve tarımsal üretime ilişkin alt yapıyı iyileştirmek, tesis etmek ve köy hayatının

## **Sivas Erzincan Kalkınma Projesi (SEKP) Verilerinin Veri madenciliği ile Sınıflandırılması ve Kümelenmesi**

kalitesini geliştirmek, proje illeri için önem arz eden ürünlerin pazar entegrasyonunu sağlamak, eğitim ve yayım yoluyla; yerel kalkınma kapasitelerini geliştirmek için 2005 yılında başlatılan bir projedir (<http://www.tarim.gov.tr/TRGM/>

[Sayfalar/DuyurularDetay.aspx?rid=97&ListName=Duyurular&refId](http://www.tarim.gov.tr/TRGM/Sayfalar/DuyurularDetay.aspx?rid=97&ListName=Duyurular&refId) Erişim Tarihi: 25.12.2013).

Çalışmada üç bölümden oluşmaktadır. Birinci bölümde veri madenciliği, sınıflandırma ve kümeleme teknikleri hakkında bilgi verilmiştir. İkinci bölümde konuyla ilgili literatür çalışması yapılmıştır. Üçüncü bölümde ise mevcut veriler üzerine sınıflandırma ve kümeleme algoritmaları uygulanmış ve sonuçlar değerlendirilmiştir.

## **II. Veri Madenciliği, Sınıflandırma ve Kümeleme**

Veri madenciliği veritabanı sistemleri, istatistik, makine öğrenmesi, görselleştirme ve bilişim bilimini kapsayan disiplinlerarası bir alan, bir disiplinler kümesinin bileşimidir (Han and Kamber, 2006:29).

Veri madenciliği, daha önceden bilinmeyen örüntüleri keşfetmek için veri keşfetme bilimi ve teknolojisi olarak veri tabanlarında bilgi keşfetmeye yönelik genel sürecin bir parçasıdır. Günümüzün bilgisayar odaklı dünyasında, bu veri tabanları büyük çapta bilgiyi içerirler. Bu bilgilerin erişilebilirliği ve bolluğu veri madenciliğini çok önemli ve gerekli hale getirmektedir (Rokach and Maimon, 2008:1).

Veri madenciliği farklı bakış açılarından veri analiz etmeye dayalı bilgi keşfetmek ve reklam, bioenformatik, veri tabanı pazarlaması, sahteciliğin

tespiti, e-ticaret, sađlık, gvenlik, web, finansal tahmin vb. dahil olmak zere eřitli uygulamalara tatbik edilebilir faydalı bilgiler retmek olarak bilinir ( Jain, 2011).

Gartner Grubuna gre "Veri madenciliđi rnt tanıma teknolojilerinin yanısıra istatistiki ve matematiksel teknikleri kullanarak depolarda saklanan byk aptaki verileri szme suretiyle anlamlı yeni iliřkiler, rntler ve trendler keřfetme srecidir" (Larose, 2005:2).

En yaygın veri madenciliđi algoritmaları ve modelleri -ki bunlar karar ađaaları, birliktelik kralları, kmeleme, sınıflandırma, oklu dođrusal regresyon, ardısıřık rntler ve zaman serisi tahminlerini kapsamaktadır- kuraklık rntleri ve karakteristiklerini teřhis etme potansiyeline sahiptir. Birliktelik kuralları, kmeleme ve ardısıřık keřif yaklařımları kuraklıđın oluřumunu ve yođunluđunu arařtırmak ve tahmin etmek iin faydalı aralar olabilir. Buna karřılık sınıflandırma, regresyon ve zaman serisi analizleri kuraklık rntlerinin haritalandırılması ve izlenmesi aısından uygun olabilir (Tadesse 2009).

Sınıflandırma eřitli nesnelerin farklı sınıflara ayrısıřtırılması iřlemidir. Bu sınıflar iřletme kuralları, sınıf sınırları veya bir matematiksel fonksiyon ile tanımlanabilir. Sınıflandırma iřlemi sınıflandırılacak varlıđın karakteristikleri ile bilinen bir sınıf tayini arasındaki iliřkiye dayalı olabilir. Bu tarz bir sınıflandırma denetimli diye bilinir. Eđer bir sınıf iin bilinen rnekler mevcut deđilse, sınıflandırma denetimsiz olur. En yaygın denetimsiz sınıflandırma yaklařımı kmelemedir. Kmeleme teknolojisinin en yaygın uygulamaları pazar sepeti analizi dahil perakende rn benzeřme analizinde ve sahtecilik tespitindedir (Nisbet, et al., 2009: 235).

Birliktelik kuralları, kmeleme, karar ađaaları, diskriminant analizi, yapay sinir ađları, genetik algoritmalar vb. gibi ok sayıda veri madenciliđi algoritması bulunmaktadır. Bu algoritmalar bir yneticinin kararlarını ynlendirebilecek

### **Sivas Erzincan Kalkınma Projesi (SEKP) Verilerinin Veri madenciliği ile Sınıflandırılması ve Kümelenmesi**

bilgileri keşfetme ve bilgi sağlama amacıyla çeşitli alanlardan bilgi işleme amacıyla kullanılabilir. Enformasyon geçmişle ve bugünle ilişkili verilerdir. Bilgi orijinal verilere dayalı gelecek trendlerin ve orijinal verilerden çıkarılacak gerekli enformasyon için bir temel sağlar. Açıkçası bilgi ve enformasyon veriler yoluyla ilişkilendirilmektedir (Wu and Li 2003).

Küme analizi, bir gruptaki örneklerin benzer olması ve farklı gruplara ait örneklerin benzer olmaması için, bir ilişkilendirme ölçüsü kullanarak, örneklerin bir dizi grup içerisine otomatik olarak sınıflandırılması yönünde bir yöntemler setidir. Bir küme analizi sisteminin girdisi iki örnek arasındaki bir örnekler dizisi ve bir benzerlik (ya da benzersizlik) ölçüsüdür. Küme analizinin çıktısı bir veri setinin bir bölüntüsünü ya da bir bölüntüler yapısını oluşturan bir dizi gruptur (kümedir). Küme analizinden gelen diğer bir sonuç her kümenin genel bir tarifidir, ve bu, özellikle veri setinin karakteristiklerinin daha derin bir analizi için önemlidir (Kantardzic 2003:117).

Bayesyen sınıflayıcılar istatistiki sınıflayıcılardır. Onlar muayyen bir veri grubunun belirli bir sınıfa ait olma ihtimali gibi sınıf üyeliği ihtimallerini tahmin edebilirler (Han and Kamber 2006:310).

Bayesyen yöntem bu dışsal enformasyonu veri analizi sürecine katmak için ilkeli bir yol sağlar. Bu süreç analiz edilmiş veri seti için hâlihazırda muayyen bir ihtimal dağılımıyla başlar. Bu dağılım, herhangi bir veri değerlendirilmeden önce muayyen olduğu için, önsel bir dağılım olarak adlandırılır. Yeni veri seti bu önsel dağılımı arka bir dağılıma günceller (Kantardzic 2003:96).

Bayesyen literatür bu meseleyle ilgili iki genel strateji getirir. *Model seçimi*, tek bir "en iyi" modeli seçer ve sonraki tahminleri bu modele dayandırır. *Model Ortalaması*, modelleri birleştirir ve tahminleri ağırlıklı ortalama olarak hesaplar. Her iki yaklaşımın da güçlü ve zayıf yönleri vardır. Model seçimi

hesaplama açısından daha kolaydır ve incelemek ve eleştirmek için tek bir model getirir. Bununla birlikte tek bir modele göre belirlenen tahminler model belirsizliğini göz ardı etmekte ve kötü bir şekilde ölçülebilmektedir. Model ortalaması hesaplama itibariyle zor ve anlaşılmaz olabilir, fakat ekseriyetle daha iyi tahminler vermektedir (Madigan and Ridgeway 2003).

### III. Literatür Özeti

Konuyla ilgili literatürde farklı veri setleri üzerinden yapılmış çok sayıda çalışma bulunmaktadır. Bunlardan Olgun ve Özdemir (2012), kontrol grafikleri üzerinde Yapay Sinir Ağları ve Bayes sınıflandırıcıların doğru sınıflandırma performanslarını karşılaştırmışlar ve Bayes sınıflandırıcının performansının daha yüksek olduğu sonucuna varmışlardır.

Pandey et al (2011), yükseköğretim öğrencilerine ait verileri bayesyen sınıflayıcı ile sınıflandırarak uygun zamanda uygun eylem ile öğrenci başarısızlığının azaltılabileceğini ortaya koymuşlardır.

Palaniappan and Awang (2008), Intelligent Heart Disease Prediction System (IHDPDS) sistemine ait verileri kullanarak, Karar ağaçları, Naive Bayes ve Yapay sinir ağları ile sınıflandırma yapmışlar ve kalp hastalarının teşhisinde bu sonuçların hemşire ve tıp öğrencilerine yardım edebileceğini belirtmişlerdir.

Bhargavi P. & Jyothi S. (2009), tarımsal toprak verilerini kullanarak WEKA yazılımında veri madenciliği yapmış ve verileri sınıflandırmalardır. Analiz sonuçlarına göre Naive Bayes sınıflayıcı %100 doğru sınıflandırma oranıyla en başarılı algoritma olmuştur. Yazarlar bu araştırmalarıyla veri madenciliğinin gelecekteki toprak araştırmalarında iyi bir araç olabileceğini göstermişlerdir.

Frank et al (2004), Bioinformatik alanında WEKA yazılımının kullanımını anlatmışlar, karar ağaçları, kural kümeleri, bayes sınıflayıcılar, SVM (Support

#### **Sivas Erzincan Kalkınma Projesi (SEKP) Verilerinin Veri madenciliği ile Sınıflandırılması ve Kümelenmesi**

Vector Machines), lojistik ve lineer regresyon, MLP (Multi-Layer Perceptron) ve en yakın komşu gibi önemli sınıflandırma ve regresyon tekniklerini desteklediğini ve k-means ve EM gibi kümeleme algoritmalarını desteğine vurgu yapmışlardır.

Ngai et al ((2009), belli bir aralıktaki Müşteri ilişkileri yönetimi (CRM) ve veri madenciliği makalelerini tarayarak sınıflandırmışlar ve veri madenciliğinin müşteri ilişkilerinde kullanımını irdelemişlerdir.

Abraham et al (2007), medikal veri madenciliği için kendilerinin geliştirmiş olduğu bayes tabanlı bir sınıflayıcı ile yaptıkları madencilik sonucunda bayes sınıflandırıcının performansının lojistik regresyon ve Support Vector Machines (SVM) algoritmasına göre daha iyi olduğunu ortaya koymuşlardır.

Kirkos et al (2007), veri madenciliği algoritmalarından karar ağaçları, Yapay Sinir Ağları ve Bayes Ağlarını kullanarak hileli finansal tabloların tespit edilebileceğini ortaya koymuşlardır.

DIMÍĆ et al. (2011), Moodle elektronik öğrenme materyallerini kullanarak öğrenci verilerini toplamış ve bu data ile hem sınıflandırma, hem kümeleme, hem de birliktelik kuralları teknikleri ile analiz yapmışlardır.

#### **IV. Veri Seti ve Yöntem**

Bu çalışmada Sivas Erzincan Kalkınma Projesi (SEKP)'ne ait veriler kullanılmıştır. SEKP kapsamında 762 başvuruya ait 6096 veri girişi yapılmıştır. Sivas Tarım İl Müdürlüğünden temin edilen veriler Excel formatında alınmıştır. Excel makroları kullanılarak veriler düzenlenmiş ve veri ambarı hazırlanmıştır. Gerekli değişken tanımları yapıldıktan

sonra veriler “sekp.arff” adlı metin dosyasına kaydedilmiştir. Bu verilere göre sınıflandırma ve kümeleme analizleri yapılmıştır.

Dosyada tanımlanan değişkenler aşağıdaki gibidir:

```
@relation sekp
@attribute BS {B,T}
@attribute BBHS Numeric
@attribute TBUD {E,H}
@attribute CMYBA Numeric
@attribute CKSAV Numeric
@attribute KKYD {E,H}
@attribute SEKPHD{0,1,2}
@attribute Class {Kabul,Red}
@DATA
B,37,E,43829,88463,E,0,Kabul
B,36,H,93161,308917,E,0,Kabul
B,23,H,79379,79379,E,0,Kabul
```

...

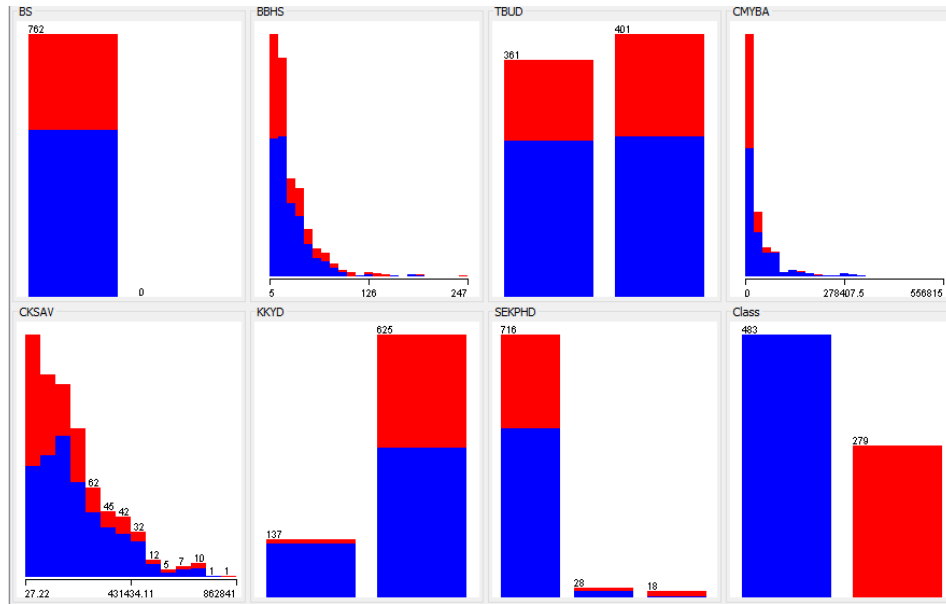
Değişkenlerde; başvuru sahibi BS olarak tanımlanmış ve “B/T” değerlerini almaktadır. “B” Bireysel başvuruyu, “T” ise “Tüzel” başvuruyu göstermektedir. Büyükbaş hayvan sayısı BBHS değişkeniyle ve sayısal (numeric) tipte tanımlanmıştır. Tarımsal birliklere üyelik durumu TBUD şeklinde tanımlanmıştır ve “E/H” değerlerini almaktadır. Çayır mera yem bitkisi alanı CMYBA değişkeniyle ve sayısal (numeric)



## Sivas Erzincan Kalkınma Projesi (SEKP) Verilerinin Veri madenciliği ile Sınıflandırılması ve Kümelenmesi

tipte tanımlanmıştır. ÇKS arazi varlığı CKSAV şeklinde ve sayısal (numeric) tipte tanımlanmıştır. Kırsak Kalkınma Yatırımlarını Destekleme Programından yararlanma durumu KKYPD şeklinde tanımlanmıştır ve “E/H” değerlerini almaktadır. SEKP hibe durumu SEKPHD şeklinde tanımlanmış ve “0/1/2” değerlerini almaktadır. “0” hiç yararlanmadı, “1” bir kez yararlandı ve “2” ise birden fazla yararlandı anlamına gelmektedir. Sınıflar ise “Kabul” ve “Red” biçiminde tanımlanmıştır. Herhangi bir değere ait veri olmaması ya da belirsizlik olması durumunda ise “?” tanımı yapılmıştır.

Aşağıda herbir değişkenin verilerine göre grafiği sunulmuştur:



Şekil 1. Değişkenlerin Grafikleri

## V. Uygulama

Yapılan çalışmada Waikato Üniversitesinde geliştirilmiş olan WEKA Programının (Waikato Environment for Knowledge Analysis) 3.6.9 sürümü kullanılmıştır. WEKA Programı, açık kaynak kodlu bir yazılımdır. Bu program pek çok sınıflandırma, kümeleme ve birliktelik kurallarına ait algoritmayı desteklemektedir. WEKA, metin tabanlı arff, arff.gz, names, data, csv, c45, libsvm, dat, bsi, xrf, xrf.gz dosya tiplerinin yanı sıra, veritabanlarını ve verilerin olduğu URL adreslerini de desteklemektedir.

**Tablo 1.** Sınıflandırma Analizine Ait Bazı Algoritmaların Başarım Dereceleri

Algoritmalar	Doğru sınıflandırılan Örnek	Kappa İstatistiği	Ortalama Mutlak Hata	Ortalama Hata Karekök	Görelü Mutlak Hata %	Görelü Hata Karekök %	TP Oran 1	FP Oran 1	F-Ölçütü
BayesNet	549	0.3814	0.3641	0.4293	78.4163	89.1037	0.72	0.349	0.716
LMT	541	0.3699	0.3609	0.4273	77.7256	88.707	0.71	0.343	0.709
ADTree	541	0.348	0.3891	0.4318	83.8188	89.6243	0.71	0.376	0.702
RBFNetwork	540	0.3607	0.3672	0.4308	79.0907	89.4249	0.709	0.354	0.706
JRip	540	0.3587	0.396	0.4498	85.3055	93.3742	0.709	0.358	0.705
NBTree	540	0.3558	0.3658	0.4331	78.7814	89.8942	0.709	0.362	0.704
Decorate	539	0.3564	0.4302	0.4477	92.6544	92.9318	0.707	0.358	0.704
DTNB	538	0.3619	0.3681	0.4307	79.2769	89.4052	0.70	0.34	0.705

**Sivas Erzincan Kalkınma Projesi (SEKP) Verilerinin Veri madenciliği ile Sınıflandırılması ve Kümelenmesi**

							6	7	
MultilayerPerceptron	537	0.3526	0.3418	0.4293	73.614	89.1069	0.705	0.358	0.702
REPTree	536	0.3369	0.3598	0.4446	77.5018	92.2875	0.703	0.379	0.696
Bagging	532	0.3273	0.3515	0.4311	75.7116	89.4792	0.698	0.382	0.692
Decision Table	530	0.3246	0.3824	0.4389	82.3566	91.1133	0.696	0.38	0.69
LWL	528	0.3444	0.3764	0.4391	81.0731	91.1433	0.693	0.345	0.694
J48	527	0.3154	0.3538	0.4448	76.207	92.3371	0.692	0.386	0.686
SimpleCart	525	0.2975	0.3793	0.4477	81.6947	92.9403	0.689	0.405	0.679
LADTree	522	0.3235	0.3517	0.4845	75.7587	100.5701	0.685	0.361	0.686
PART	519	0.2954	0.3513	0.4429	75.6744	91.9416	0.681	0.393	0.676
BFTree	516	0.3138	0.3748	0.4496	80.7266	93.3333	0.677	0.359	0.679
Ridor	513	0.225	0.3268	0.5716	70.3837	118.6582	0.673	0.469	0.646
SMO	512	0.1958	0.3281	0.5728	70.6663	118.8962	0.672	0.5	0.629
NaiveBayes	510	0.3567	0.3768	0.4407	81.1593	91.4706	0.669	0.274	0.674
Dagging	496	0.0601	0.3493	0.5004	75.2455	103.8802	0.651	0.603	0.532
RandomForest	494	0.2024	0.3517	0.4961	75.7501	102.9852	0.648	0.456	0.636
HyperPipes	486	0.0155	0.4995	0.4996	107.5976	103.6985	0.638	0.626	0.503
IB1	481	0.1814	0.3688	0.6073	79.4289	126.0524	0.631	0.455	0.625
OneR	477	0.1416	0.374	0.6116	80.5596	126.9464	0.626	0.493	0.609
VFI	469	0.2646	0.4839	0.4862	104.2264	100.9215	0.615	0.317	0.62
IBk	466	0.1205	0.3704	0.5956	79.7891	123.627	0.61	0.49	0.598

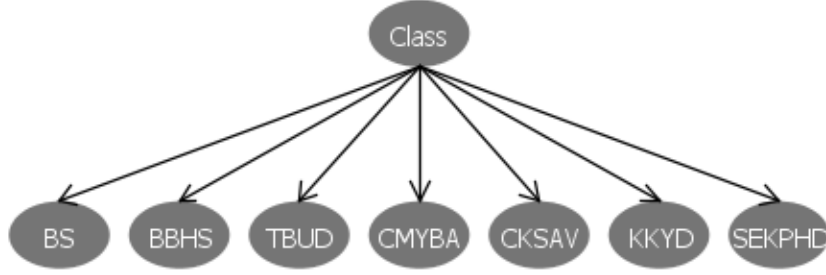
						3	2	7	
KStar	445	0.0808	0.4235	0.6286	91.2114	130.480	0.58	0.50	0.578
						6	4	5	

Tablo 1'deki sonuçlar WEKA programı ile elde edilmiş sonuçlardır. Yapılan uygulama çalışmasında program tarafından desteklenen bütün algoritmalar denenmiş ve başarımlarına göre tabloda sıralanmıştır. Bu çalışmada BayesNet algoritması 549 doğru sınıflandırma sayısı ile en başarılı algoritma olmuştur. Bu algoritmanın kappa istatistiği 0.3814, True Positive oranı 0.72, False Positive oranı ise 0.349 ve F-ölçütü ise 0.716 olarak gerçekleşmiştir. Bu algoritmayı 541 doğru sınıflandırma sayısı ile LMT ve ADTree algoritmaları gelmektedir. Daha sonra diğer algoritmalar gelmektedir.

BayesNet algoritmasının sınıfları olan Kabul seçeneğinin ihtimal dağılımı 0.634, Red seçeneğinin ihtimal dağılımı ise 0.366'dır. Herbir değişken içinde ihtimal dağılımları verilebilir.

BayesNet algoritmasının görsel grafiği ise aşağıdaki gibidir:

**Sivas Erzincan Kalkınma Projesi (SEKP) Verilerinin Veri madenciliği ile Sınıflandırılması ve Kümelenmesi**



**Şekil 2.** *BayesNet Görsel Grafiği*

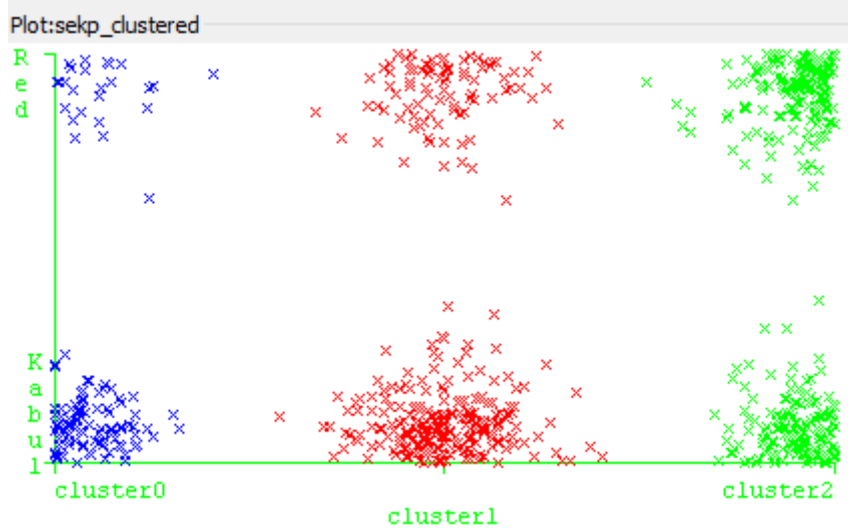
Mevcut verilerin EM algoritmasıyla yapılan kümeleme analizi sonucunda ise 3 küme oluşmuştur. Tam eğitilmiş set kullanılarak yapılan analiz sonucunda oluşan 0 numaralı birinci kümede 122 proje bulunmakta ve %16 sına denk gelmektedir. 1 numaralı ikinci kümede ise 318 proje bulunmaktadır ve %42 sine denk gelmektedir. 2 numaralı üçüncü kümede ise 322 proje bulunmakta ve %42 sine denk gelmektedir. Oluşan kümeler aşağıdaki gibidir:

**Clustered Instances**

0	122 ( 16%)
1	318 ( 42%)
2	322 ( 42%)

**Şekil 3.** *EM algoritmasıyla oluşan kümeler*

WEKA kümeleyicisinin EM algoritması sonucunda oluşan küme görseli aşağıdaki gibidir:



Şekil 4. EM kümeleme görseli

## VI. Sonuç

Bu çalışmada, sınıflandırma ve kümeleme yöntemi ile SEKP verileri kullanılarak analiz yapılmıştır. Veri madenciliği, gizli, önemli, önceden bilinmeyen, yararlı bilgileri ortaya koyan bir veri analiz tekniğidir. Bu yöntemle, alışlagelmiş analiz tekniklerinden farklı olarak, yalnızca sayısal verilerle değil, sayısal olmayan veriler ile de analizler yapılabilmekte ve gizli örüntüler ortaya çıkarılabilmektedir. Veri madenciliği analizinde kullanılan algoritmaları destekleyen pek çok program geliştirilmiştir. Bu alanda yaygın olarak kullanılan

programlardan biri de WEKA programıdır. WEKA programı hemen hemen bütün veri madenciliği yöntemlerini desteklemekte ve her birine ait pek çok algoritma ile analizler yapabilmektedir. Ayrıca WEKA programı görsel olarak ta yararlı sonuçlar üretebilmektedir. Söz konusu programı kullanarak yaptığımız uygulama sonucunda veri ambarındaki verileri en iyi sınıflandıran algoritmanın BayesNet algoritması olduğu ortaya çıkmıştır. Ayrıca EM algoritmasıyla üç küme üretilmiştir.

## VII. KAYNAKÇA

- [1] Abraham, Ranjit & Simha, Jay B.& Iyengar S.S (2007), **Medical Datamining with a new algorithm for Feature Selection and Naïve Bayesian Classifier**, 10th International Conference on Information Technology, DOI 10.1109/ICIT.2007.41, IEEE Computer Society, pp.44-49
- [2] Bhargavi, P. & Jyothi, S. (2009), **Applying Naive Bayes Data Mining Technique for Classification of Agricultural Land Soils**, IJCSNS International Journal of Computer Science and Network Security, Vol.9 No.8, pp. 117-122
- [3] Dimić, G. & Kuk, K. & Ahorjanski, M. (2011), **Mining Student's Data For Analyze Electronic Learning Materials Available On The Moodle Course** Metalurgia International Vol. XVI no. 12, 2011, pp.78-82
- [4] Frank, Eibe & Hall Mark & Trigg, Len & Holmes Geoffrey & Witten H. (2004), **Data mining in bioinformatics using WEKA**, Bioinformatics, Vol. 20 no. 15 2004, pp.2479–2481
- [5] Giudici, Paolo and Figini, Silvia, (2009) **Applied Data Mining For Business and Industry**, Second Edition, Wiley Publication, West Sussex, pp.90-91
- [6] Han, Jiawei and Kamber, Micheline, (2006), **Data Mining: Concepts and Techniques, Second Edition**, Morgan Kaufmann Publications, San Francisco
- [7]<http://www.tarim.gov.tr/TRGM/Sayfalar/DuyurularDetay.aspx?rid=97&ListName=Duyurular&refId>
- [8] Jain, Y. K., Yadav, V. K., Panday, G. S., (2011), **An Efficient Association Rule Hiding Algorithm For Privacy Preserving Data Mining**, *International Journal On Computer Science And Engineering*, Vol. 3 No. 7, pp. 2792-2798.
- [9] Kantardzic, Mehmed , (2003). **Data Mining: Concepts, Models, Methods, and Algorithms**, John Wiley & Sons J. B. Speed Scientific School, University of Louisville IEEE Computer Society, Sponser.
- [10] Kirkos, Efstathios & Spathis, Charalambos & Manolopoulos, Yannis (2007), **Data Mining techniques for the detection of fraudulent financial statements**, *Expert Systems with Applications* 32, pp. 995–1003

- [11] Larose, Daniel T., (2005). **Discovering Knowledge In Data**, Wiley Publication, New Jersey.
- [12] Madigan David and Ridgeway Greg (2003). "Bayesian Data Analysis", **The Handbook of Data Mining**, Edited by. Nong Ye, New Jersey: Lawrence Erlbaum Associates, Inc. pp.103-131
- [13] Ngai, E.W.T.& Xiu, Li & Chau, D.C.K. (2009), **Application of data mining techniques in customer relationship management: A literature review and classification**, Expert Systems with Applications 36 pp. 2592–2602
- [14] Nisbet, R., Elder, J., and Miner, G., (2009). **Handbook of Statistical Analysis and Data Mining Applications**, Elsevier Inc, Burlington.
- [15] Olgun Mehmet Onur & Özdemir Gültekin (2012), **İstatistiksel Özellik Temelli Bayes Sınıflandırıcı Kullanarak Kontrol Grafiklerinde Örüntü Tanıma**, Gazi Üniv. Müh. Mim. Fak. Der. Cilt 27, No 2, pp. 303-311
- [16] Palaniappan, Sellappan & Awang, Rafiah (2008), **Intelligent Heart Disease Prediction System Using Data Mining Techniques**, IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.8, pp. 343-350
- [17] Pandey, Umesh Kumar & Pal, S. (2011), **Data Mining : A Prediction Of Performer Or Underperformer Using Classification**, (IJCSIT) International Journal Of Computer Science And Information Technologies, Vol. 2 (2) , pp. 686-690
- [18] Rokach, Lior and Maimon Oded (2008), **Data Mining With Decision Trees**, World Scientific, New Jersey
- [19] Tadesse, T., Wardlow, B. And Hayes, M.J. (2009), "The Application of Data Mining for Drought Monitoring and Prediction" , **Data Mining Applications for Empowering Knowledge Societies**, Ed. Hakikur Rahman, Hershey • New York, pp.280-291
- [20] Wu, Tong and Li, Xiangyang (2003). "Data Storage and Management", **The Handbook of Data Mining**, Edited by. Nong Ye, New Jersey: Lawrence Erlbaum Associates, Inc. pp.393-407
- [21] Weiss, Sholom M. And Zhang, Tong (2003), Performance Analysis And Evaluation, **The Handbook of Data Mining**, Edited by Nong Ye *Arizona State University*, Lawrence Erlbaum Associates, Mahwah, pp.425-440