

TELE PAZARLAMA VERİLERİNİN VERİ MADENCİLİĞİ TEKNİKLERİ KULLANILARAK SINIFLANDIRILIP ANALİZ EDİLMESİ

Murat ALTUNTAŞ

Gebze Teknik Üniversitesi

Bilgisayar Mühendisliği

E-mail: murataltuntas44@gmail.com

ÖZET

Yapılan bu projede Tele Pazarlama konusunu işledim. Veri madenciliği ile pazarlama arasındaki yoğun etkileşim günümüz işletmelerinde yaygın bir hal almıştır. Doğrudan pazarlamanın bir alt kolu olarak tanımlanabilecek tele pazarlama bankalar tarafından yoğun olarak kullanılmaktadır. Ancak bankaların müşterileriyle yaptığı görüşmeler yoğunlukla müşteriler tarafından hoş karşılanmamakta hatta yoğunlukla yapılan çağrılara müşteriler hiç yanıt vermemeyi tercih etmektedirler. Tele pazarlama başarımındaki bu düşüşü arttırabilmek için veri madenciliği önemli bir araçtır. Bankalar müşterilerini daha iyi tanıdıkça onlarla daha sağlıklı ilişki kurabilecektir. Bu çalışmada da veri madenciliği yöntemlerinden Knn, SVM ve Karar Ağaçları yöntemleri ile yapılan analizler sonucunda bir bankanın müşterileri hakkında aydınlatıcı bilgiler üretilmeye çalışılmıştır. Bu çalışmada Portekiz'de bulunan bir bankanın verilerini analiz edilmiştir. Bankaların veya müşterileriyle tele pazarlama yoluyla iletişim kuran diğer firmaların müşterilerini daha iyi

tanımalarına ve tekliflerini en uygun yolla belirli müşterilere iletmelerine yönelik örnek bir veri madenciliği projesi yapmayı hedefledim.

Anahtar Kelimeler: Veri madenciliği, doğrudan pazarlama, tele pazarlama.

ABSTRACT

The intense interaction between marketing and data mining is common today's businesses. A new approach created called; database marketing. Business managers began to use this approach more widely especially in terms of customer segmentation and Customer Relationship Management. In this work, we developed a data mining application-which is rapidly growing and becoming limelight- for direct marketing. As a sub-branch of direct marketing, telemarketing is used often by banks. But today, customers do not enjoy this and commonly they prefer not to answer calls. To prevent from this decline in the popularity of telemarketing, data mining is an important tool. Banks can understand customers and their behaviors more deeply using such application. As a result, they are able to develop a stronger relationship with

them. In this work, we tried to produce some beneficial information about bank customers via the results of association rules. It offers a case data mining study for banks or other businesses which employ telemarketing for know their customers better and convey their proposals to specific costumers in most appropriate way.

Key Words: Data mining, direct marketing, telemarketing.

GİRİŞ

Veri madenciliği gelişen veri tabanı teknolojisine paralel olarak çok hızlı bir gelişme göstermiştir. Kullanılan algoritmalar, bilgisayar yazılımları, uygulama alanları çeşitlenmiştir. Çevrimiçi alışverişte müşterilere alışveriş yaparken geçmişte yaptığı alışverişler analiz edilerek anında yeni teklifler sunulur hale gelmiştir. Sigorta ve finans şirketleri veri madenciliği ile müşterilerin riskliliği, dolandırıcılık tespiti (fraud detection) gibi işlemleri gerçekleştirmektedir. Veri madenciliğinin yaygın kullanım alanlarından birisini ise pazarlama teşkil etmektedir. İlgili yayınlar incelendiğinde pazarlama ile veri madenciliği etkileşimine çok sık rastlamak mümkündür.

TELE PAZARLAMA

Tele pazarlama potansiyel müşterileri çekme, mevcut müşterilere satış yapabilme ve müşterilerin sorularına yanıt vererek veya sipariş alarak hizmet sağlama amacıyla telefonların ve çağrı merkezlerinin kullanımı olarak ifade edilebilir. Tele pazarlama genel olarak iki yönlüdür. Birincisi firmadan müşterilere hizmet ve teklif sunma amaçlı gerçekleştirilen dışa doğru tele pazarlamadır. İkincisi ise müşterilerin istek ve sorunlarını yanıtlamak suretiyle hizmet

vermek veya ürün veya hizmet satın alabilmek amacıyla gerçekleştirdiği içe doğru tele pazarlamadır.

Tele pazarlama yöntemini kullanabilmek için işletmeler genelde bir çağrı merkezine ihtiyaç duyarlar. Bir çağrı merkezi bilgisayarlarla desteklenmiş çalışanların, içe doğru aldıkları veya dışa doğru yaptıkları telefon çağrılarının, otomatik çağrı dağıtım sistemi veya tahmine dayalı çevirme sistemleriyle kontrol edildiği ve işleme sokulduğu bir işlemler bütünüdür.

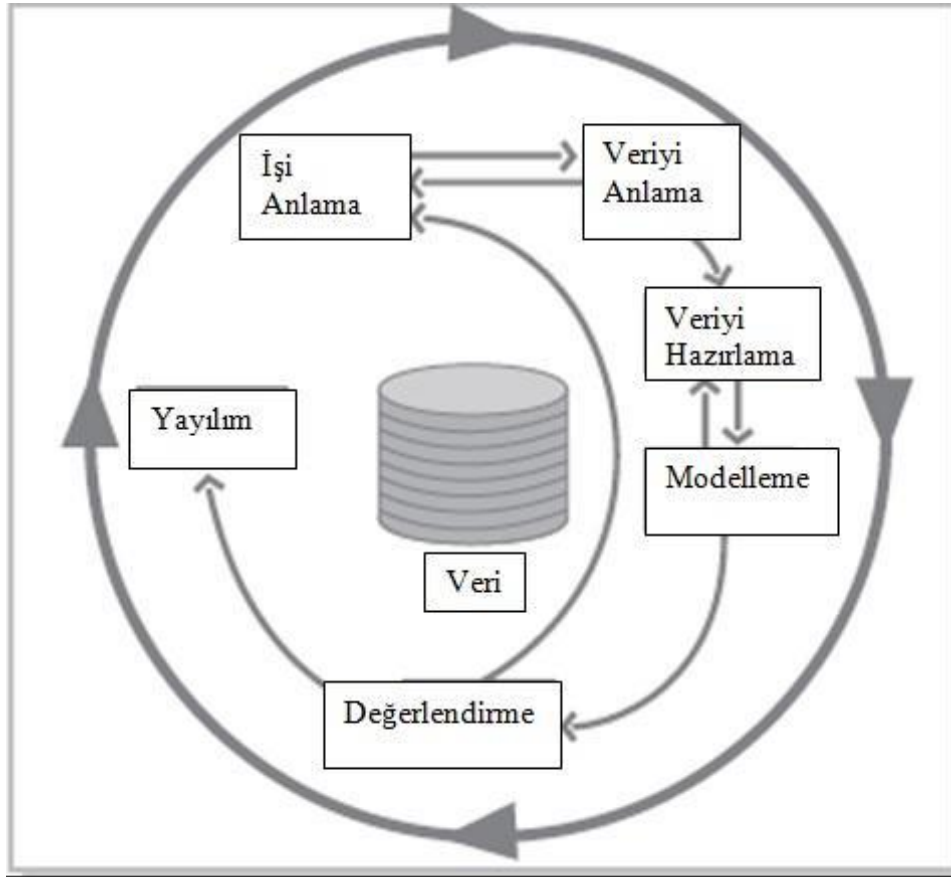
UYGULAMA

Uygulama da veri madenciliği süreci aşağıdaki basamakları takip etmektedir:

- İşi kavrama
- Veriyi kavrama
- Veriyi hazırlama
- Modelleme
- Değerlendirme
- Yayılım

İşi Kavrama

Son dönemde müşteriler istedikleri ürün veya hizmete çok hızlı bir şekilde ulaşabildikleri gibi ürün ve hizmetler hakkında bilgilere de kolayca ulaşabilmektedir. Dolayısıyla kendilerine ulaştırılan bu çağrılar artık rahatsız edici bir hal almıştır. Hem rahatsız edici yapısı hem de yüksek maliyeti düşünülerek tele pazarlama yoluyla ulaşılan müşteri kitlesini azaltmak ve başarıyı arttırmak burada esas alınmıştır. Müşterilerin nitelikleri ve kampanya teklifi arasındaki bilinmeyen gizli ve bankaya ışık tutacak bilgilerin ortaya çıkarılması ve performansı arttırmaya yarayacak doğrudan pazarlama stratejileri için ipuçları oluşturmak hedef olarak belirlenmiştir.



Veriyi Kavrama

Uygulama kapsamında kullanılan veri tabanı UCI Makine Öğrenmesi ve Akıllı Sistemler isimli veri tabanı deposundan sağlanmıştır. Kaynak birçok bilimsel çalışmada kullanılan veri setlerini barındırmakta ve Amerika Birleşik Devletleri'nde bulunan Milli Bilim Vakfı Kurumu tarafından (The National Science Foundation) desteklenmektedir. Aynı veri tabanı kullanılarak daha önce bir çalışma gerçekleştirilmiştir. Moro ve arkadaşları (2011) yılında bu veri tabanını kullanarak veriler hakkında en iyi tahmin edici modelleri tartışmış daha sonra en anlamlı nitelikleri araştırmıştır. Veri tabanındaki veriler Portekiz'deki bir bankanın doğrudan pazarlama kampanyaları hakkındadır. 17 farklı kampanya telefon görüşmeleriyle müşterilere ulaştırılmıştır. Sıklıkla bir müşteriyle teklifi ulaştırabilmek için birden fazla kez görüşüldüğü görülmektedir. Toplam 79354 telefon görüşmesi

yapılmıştır. Bu görüşmeler sırasında uzun vadeli ve iyi bir faiz oranı olan mevduat önerilmiştir. 45211 örneklem ve 16 nitelik (ve bir de sonuç niteliği) barındıran veri tabanında herhangi bir kayıp veriye rastlanmamıştır. Burada kullanılan nitelikler aşağıda açıklanmıştır:

1. Yaş: Müşterilerin yaşı sayısal olarak mevcuttur.

2. Meslek: Meslekler gruplar halinde veri tabanına girilmiştir; Üst kademe yönetici, bilinmeyen, işsiz, yönetim, müstahdem, girişimci, öğrenci, mavi yakalılar, serbest meslek erbabı, emekli, teknisyen, hizmet sektöründe çalışan şeklinde seçenekler mevcuttur.

3. Evlilik durumu: Evlilik durumu da gruplar halinde bulunmaktadır; evli, boşanmış ve bekâr. Boşanmış dul olan müşteriler için de kullanılmıştır.

4. Eğitim: Bilinmeyen, orta dereceli, ilk dereceli, üçüncü dereceli olarak gruplar halinde belirtilmiştir.

5. Yükümlülüğü yerine getirilmeyen kredinin olup olmaması: Evet ve hayır olarak ikili terim olarak ifade edilmiştir.

6. Bakiye: Yıllık ortalama bakiye sayısal olarak ve Avro (Euro) para birimine göre ifade edilmiştir.

7. Konut kredisi: Evet veya hayır olarak ikili terim olarak ifade edilmiştir.

8. Bireysel kredi: Evet veya hayır olarak ikili terim olarak ifade edilmiştir.

9. İletişim: İletişim kurma tarzı da 3 değişkenden oluşan grup halinde ifade edilmiştir; bilinmeyen, cep telefonu, sabit telefon.

10. Gün: Ay içerisinde en son görüşülen gün sayısal olarak ifade edilmiştir.

11. Ay: Yıl içerisinde iletişim kurulan en son ay; kategorik olarak 12 aydan oluşmaktadır.

12. Süre: En son iletişim süresi sayısal olarak saniye esasına göre ifade edilmiştir.

13. Kampanya: Kampanya süresince bu müşteriyle kurulan iletişim sayısı (son görüşme de dâhil olmak üzere).

14. Geçen gün: Müşteriyle bir önceki kampanya sırasında en son ulaşıldığı günden bu yana geçen gün sayısı sayısal olarak ifade edilmiştir (-1 daha önce müşteriyle iletişime geçilmediğini ifade etmektedir).

15. Önceden kurulan iletişim sayısı: Müşteriyle bu kampanyadan önce gerçekleştirilen iletişim sayısı sayısal olarak ifade edilmiştir.

16. Önceki Çıktı: Bir önceki kampanyanın başarılı olup olmadığı 4 şekilde ifade edilmiştir; bilinmeyen, diğer, başarısız, başarılı)

17. Çıktı-Hedef değişken (y) : Önerilen mevduat teklifin kabul edilip edilmediği ikili olarak evet ve hayır şeklinde ifade edilmiştir.

Veriyi Hazırlama

Bu aşama genelde en zaman alıcı basamaktır. Çalışmada kullandığımız veri tabanı düzenli bir şekilde bulunduğu için bu aşama kolayca gerçekleştirilmiştir. Veri tabanı dosyası Microsoft Excel aracılığıyla bir takım işlemlere tabii tutulmuştur. Bu işlemlerin yeterli gelmediği durumlarda, veri tabanı dosyası Microsoft Word'de çalıştırılıp bir takım düzenlemelere tabi tutulduktan sonra tekrar Microsoft Excel formatında çalıştırılıp en son şekli verilmiştir.

Modelleme

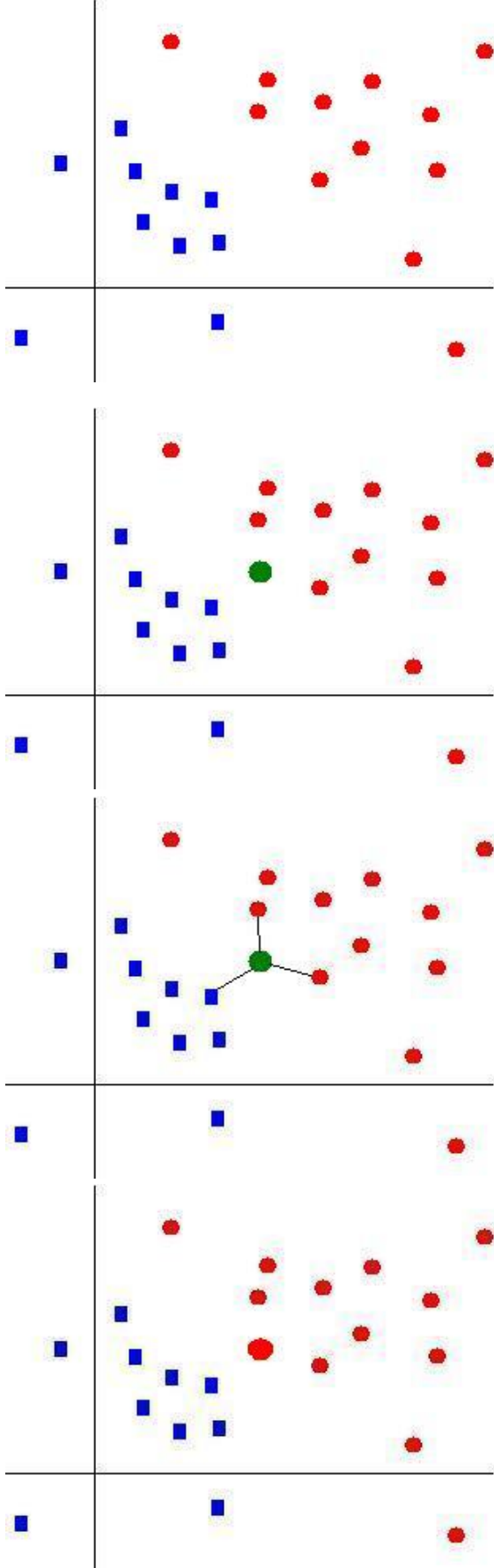
Uygulamam da 3 farklı sınıflandırma yöntemi kullanarak veri setim üzerinde çalıştım. Kullandığım sınıflandırıcılar:

- K Nearest Neighborhood (Knn)
- Support Vector Machine (SVM)
- Decision Trees

1.Knn

Sınıflandırmada kullanılan bu algoritmaya göre sınıflandırma sırasında çıkarılan özelliklerden, sınıflandırılmak istenen yeni bireyin daha önceki bireylerden k tanesine yakınlığına bakılmasıdır.

Örneğin $k = 3$ için yeni bir eleman sınıflandırılmak istensin. Bu durumda eski sınıflandırılmış elemanlardan en yakın 3 tanesi alınır. Bu elemanlar hangi sınıfa dâhilse, yeni eleman da o sınıfa dâhil edilir.



R dilini kullanarak kodladığım bu algorithmada mesafe hesabı için Manhattan mesafesi ve öklit mesafesi kullandım.

$$d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}.$$

Şekil 1 Öklit Uzaklığı

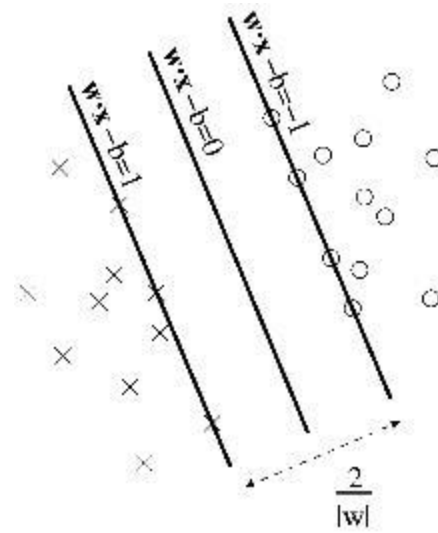
$$d_1(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|_1 = \sum_{i=1}^n |p_i - q_i|,$$

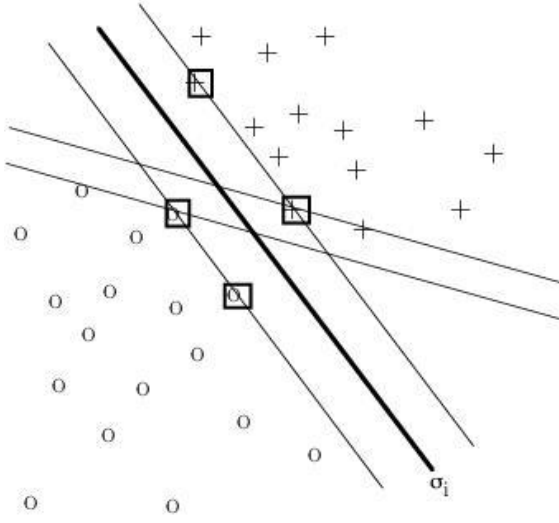
Şekil 2 Manhattan Uzaklığı

2.SVM

Sınıflandırma konusunda kullanılan oldukça etkili ve basit yöntemlerden birisidir. Sınıflandırma için bir düzlemde bulunan iki grup arasında bir sınır çizilerek iki grubu ayırmak mümkündür. Bu sınırın çizileceği yer ise iki grubun da üyelerine en uzak olan yer olmalıdır. İşte SVM bu sınırın nasıl çizileceğini belirler.

Bu işlemin yapılması için iki gruba da yakın ve birbirine paralel iki sınır çizgisi çizilir ve bu sınır çizgileri birbirine yaklaştırılarak ortak sınır çizgisi üretilir.



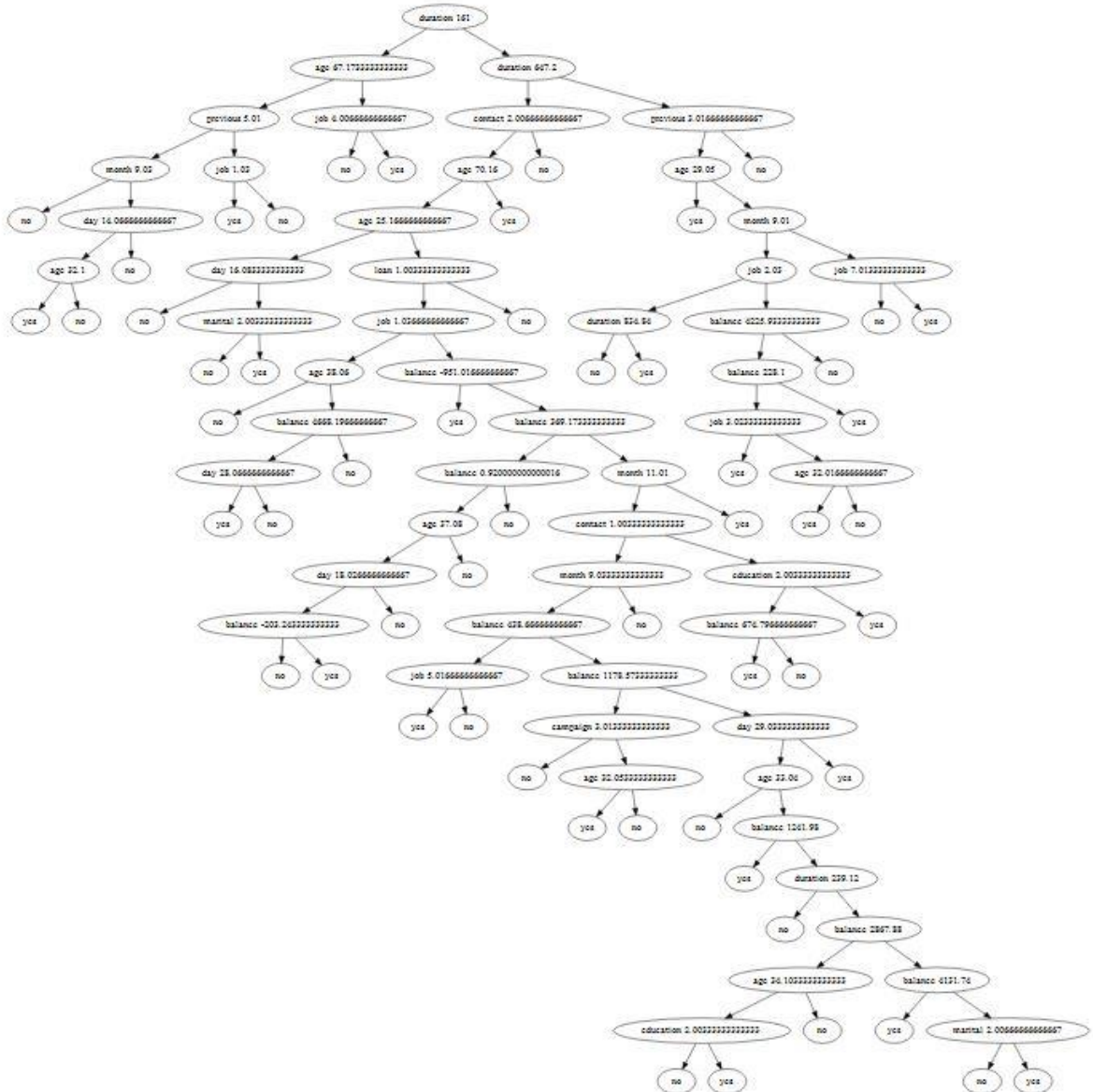


SVM'i R dilinde kullanabilmek için *library(kernlab)* kütüphanesini yükledim. İmplementasyonda *ksvm* metodunu kullandım.

İmplementasyonu, lineer svm ve polinomal svm olmak üzere iki farklı şekilde yaptım. Bu sayede veri setim üzerinde iki farklı yöntemin arasındaki farkı test ettim.

3.Decision Trees

Karar ağaçları, sınıfları bilinen örnek veriden tümevarım yöntemiyle öğrenilen ağaç şekilli bir karar yapısı çeşididir. Bir karar ağacı, basit karar verme adımları



uygulanarak, büyük miktarlardaki kayıtları, çok küçük kayıt gruplarına bölerek kullanılan bir yapıdır. Her başarılı bölme işlemiyle, sonuç gruplarının üyeleri bir diğeriyle çok daha benzer hale gelmektedir. Büyük veri tabanlarının kullanıldığı pek çok sınıflama probleminde ve karmaşık ya da hata içeren bilgilerde karar ağaçları yararlı bir çözüm olmaktadır. Tahmin edici ve tanımlayıcı özelliklere sahip olan karar ağaçları, Veri Madenciliğinde kuruluşlarının kolay olması, yorumlanmalarının kolay olması, veri tabanı sistemlerine kolayca entegre edilebilmeleri, güvenilirliklerinin daha iyi olması nedenleri ile sınıflama modelleri içerisinde en yaygın kullanıma sahip olan bir tekniktir.

Projemde karar ağaçlarını implement ederken information gains hesabı yöntemini kullanarak yaptım. Ayrıca veri kümem üzerinde k-cross validation ve pruning yöntemlerini de uyguladım.

Değerlendirme

İkincil kaynaktan sağlanan veriler kullanarak başlatılan bu veri madenciliği süreci bankaların tele pazarlama konusundaki başarımını arttırmayı hedeflemiştir. Bu doğrultuda veri madenciliği yöntemlerinden knn, svm ve karar ağaçları yöntemleri kullanılarak gerçekleştirilecek analizlerin kampanya başarısının esas alınarak gerçekleştirilmesi uygun görülmüştür.

Yayılım

Süreç gerekli veri düzeltme işlemlerini gerçekleştirmek üzere ve farklı algoritmaların denenmesi suretiyle devam ettirilebilir. Ayrıca uygulanan yöntemler göz önüne alınarak bazı sınıflandırma ve kümeleme analizlerinin gelecekte

gerçekleştirilerek modelin tazelenmesi denenebilir.

SONUÇLAR

Yaptığım uygulamanın sonuçları şu şekilde çıktı.

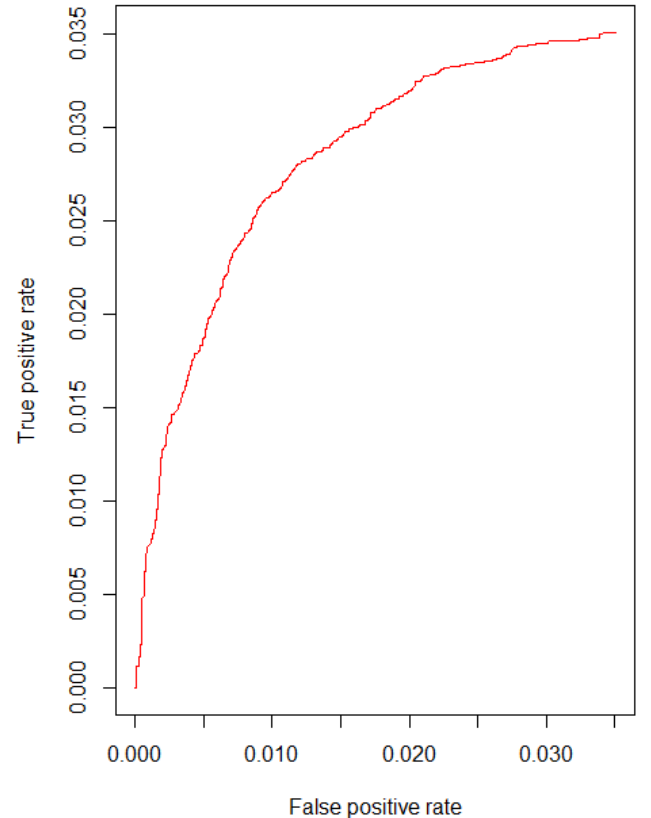
Knn algoritmasında,

- Manhattan uzaklığı formülüne göre yaptığım hesaplamada %87.35294 oranında başarılı sonuç verdi.
- Öklit uzaklığı formülüne göre yaptığım hesaplamada %90.29412 oranında başarılı sonuç verdi.

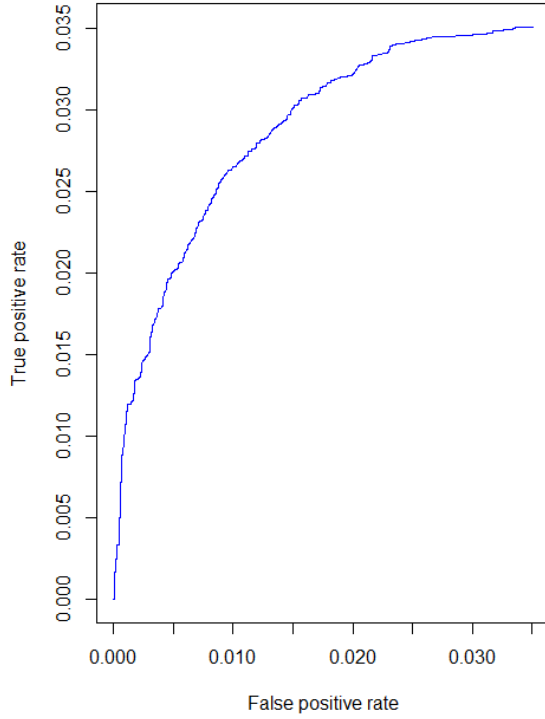
SVM algoritmasında,

- Lineer SVM'e göre yaptığım hesaplama da %86.76471 oranında başarılı sonuç verdi.
- Polinomal SVM'e göre yaptığım hesaplama da %90.29412 oranında başarılı sonuç verdi.

Lineer SVM ROC Curve sonucu:



Polinomal SVM ROC Curve sonucu:



Karar ağaçlarında kendim implement ettiğim karar ağacı algoritması dışında R dilinin kendi içinde hazır bulunan iki farklı karar ağacı fonksiyonunda da (ctree ve rpart) veri setimi test ettim. Ayrıca yaptığım işlemlerin hepsinde k-cross validation uyguladım.

Test sonuçları,

- Ctree fonksiyonuna göre yapılan hesaplamada %88.68182 oranında başarılı sonuç verdi.
- Rpart fonksiyonuna göre yapılan hesaplamada %89.07172 oranında başarılı sonuç verdi.
- Kendim implement ettiğim karar ağacına göre yapılan hesaplamada %74.15152 oranında başarılı sonuç elde ettim.

Ayrıca veri kümem üzerinde karar ağaçlarının pruning yöntemini de uyguladım. Uyguladığım pruning yönteminde entropy değerini 0.3 olarak aldım. Çıkan sonuçlar:

- Rpart fonksiyonuna göre yapılan hesaplamada %88.57576 oranında başarılı sonuç verdi.
- Kendim implement ettiğim karar ağacına göre yapılan hesaplamada %80.76869 oranında başarılı sonuç elde ettim.

KNN		
	Manhattan	%87.35294
	Öklit	%90.29412
SVM		
	Lineer	%86.76471
	Polinomal	%90.29412
Decision Tree		
	ctree	%88.68182
	rpart	%89.07172
	myDecisionTree	%74.15152
pruning -> 0.3		
	rpart	%88.57576
	myDecisionTree	%80.76869

KAYNAKLAR

Moro, Sergio, Raul Laureano, and Paulo Cortez. "Using data mining for bank direct marketing: An application of the crisp-dm methodology." (2011).

Popelka, Ondřej, et al. "Comparison of different non-statistical classification methods." *30th International Conference Mathematical Methods in Economics*. 2012.

AYTAÇ, Muhammed Bilgehan, and Hasan Şakir BİLGE. "TELE PAZARLAMA VERİLERİNİN BİRLİKTELİK KURALLARIYLA VE CRISP-DM YÖNTEMİYLE ANALİZ EDİLMESİ." *Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* 5.2 (2013): 25-39.

Hsieh, Nan-Chen. "Hybrid mining approach in the design of credit scoring

models." *Expert Systems with Applications* 28.4 (2005): 655-665.

Ngai, Eric WT, Li Xiu, and Dorothy CK Chau. "Application of data mining techniques in customer relationship management: A literature review and classification." *Expert systems with applications* 36.2 (2009): 2592-2602.

Hsieh, Nan-Chen. "An integrated data mining and behavioral scoring model for analyzing bank customers." *Expert systems with applications* 27.4 (2004): 623-633.

GELİBOLU, Levent, and Tufan ÖZSOY. "Çağrı Merkezlerinin Satış Amaçlı Kullanılması: Doğrudan Pazarlamanın Bir Unsuru Olarak Telepazarlama." *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* 22.1 (2013).

Cortez, P. 2010. "Data Mining with Neural Networks and Support Vector Machines using the R/rminer Tool". In Proceedings of the 10th Industrial Conference on Data Mining (Berlin, Germany, Jul.). Springer, LNAI 6171, 572–583.

Cortez, P. and Embrechts, M. 2011. "Opening Black Box Data Mining Models Using Sensitivity Analysis". In Proceedings of IEEE Symposium on Computational Intelligence and Data Mining (Paris, France), 341-348.

Ou, C., Liu, C., Huang, J. and Zhong, N. 2003. "On Data Mining for Direct Marketing". In Proceedings of the 9th RSFDGrC conference, 2639, 491–498.