SİGORTA ŞİRKETLERİNE YÖNELİK MÜŞTERİ SEGMENTASYONU VE MÜŞTERİ KAYIP EĞİLİMİ ANALİZ SİSTEMİ

Customer Segmentation and Customer Churn Analysis System for Insurance Companies

Tolgahan Satıcı ve Merve Bekler
Dip Danışmanlık, İstanbul, Türkiye
tolgahan.satici@dip.com.tr, merve.bekler@dip.com.tr

***Özetçe*—Müşteri bölütlemesi (customer segmentation), müşteri tabanının yaş, cinsiyet, ilgi alanları ve harcama alışkanlıkları gibi faktörler göz önüne alınarak kendi içlerinde homojen gruplara bölünmesi uygulamasıdır. Müşteri kayıp analizi (customer churn analysis) ise mevcut müşterilerin kaybını önceden tahmin etmeye yönelik bir analiz biçimidir. Çalışma kapsamında ham veriden eyleme geçirilebilir bilgi üretimi aşamalarının tümünü modüler halinde sunan bir yazılım oluşturulmuştur. Yazılım gerçek hayat problemlerinde test edilerek başarılı sonuçlar elde edilmiştir.**

Anahtar Kelimeler —müşteri kayıp analizi, segmentasyon, veri analizi, CRM.

***Abstract*—Customer segmentation is an application that segmentate customers homogeneously according to their age, gender, personal interests and spending habits. As for customer churn analysis, it is an analysis that aims predicting churn tendency of current customers beforehand. In the scope of the study, a software is developed that presents each steps of producing knowledge from raw data to actionable data as modular. Software is tested with real world problems and successful results are obtained.**

Keywords — customer churn analysis, segmentation, data analysis, CRM.

I. GİRİŞ

Teknolojik gelişmeler doğrultusunda artan veri miktarı ve bilimsel gelişmeler, veri bilimi ve makine öğrenmesi uygulamalarına ilgiyi arttırmaktadır. Uygulama alanları olarak bankacılık, e-ticaret, mobil uygulama vb. birçok iş kolunda yer alan veri bilimi, farklı alanlardaki şirketlerin ihtiyaçlarına farklı çözümler üretmektedir. Müşteri sadakatinin önem kazandığı alanlar da uygulama alanlarından biridir. Özellikle yeni müşteri elde etme maliyetinin, müşteri tutundurma maliyetlerinden 5 kat daha yüksek olduğu değerlendirildiğinde müşteri kayıp analizinin şirketler açısından önemi ortaya çıkmaktadır.

Bu araştırma kapsamında özellikle sigorta şirketlerinin başta gelen ihtiyaçlarından müşteri tanıma, ürün-müşteri eşleştirilmesi ve müşteri tutundurma konularına bilimsel yöntemler kullanarak çözüm getirecek, temelinde, esnek bir veri analiz platformu önerisi yapılmaktadır. Önerilen platformun kullanılabilirliğini ortaya koymak adına prototip uygulama geliştirilmiştir. Platform üzerinde çalışan, ilgili şirket için özelleştirilmiş bir müşteri bölütleme, müşteri kayıp analizi paketi oluşturulmuştur. Burada, araştırmamız kapsamında uygulamalara yönelik bir müşterinin kaybına sebep olabilecek değişkenler, mevcut sistem analizi ve iş bilgisi kullanılarak belirlenmiştir. Bunun yanı sıra, sigortacılık sektöründe, müşterilerin birbirlerine olan benzerliklerini belirleyen aktif değişkenler tespit edilmiştir.

Bu bildirinin organizasyon yapısı şu şekildedir: II. Bölüm'de problem tanımı ve yaklaşımlar belirtilmektedir. Müteakiben III. Bölüm'de önerilen metodoloji ve yaklaşımlar tanıtılarak IV. Bölüm'de sistem mimarisi anlatılmaktadır. V. Bölüm'de yaklaşımlarla ilgili deneysel değerlendirme ve çalışmamızın sonuçları paylaşılmaktadır. Son olarak VI. Bölüm'de sonuçlar ve gelecekteki çalışmalar aktarılmaktadır.

II. PROBLEMİN TANIMI

Müşteri bölütlemesi (customer segmentation), müşteri tabanının yaş, cinsiyet, ilgi alanları ve harcama alışkanlıkları gibi faktörler göz önüne alınarak kendi içlerinde homojen gruplara bölünmesi uygulamasıdır. Müşterilerin birçok açıdan, birbirlerinden farklı olduğundan hareketle, farklı sahalarda faaliyet gösteren her şirket için birbirine benzeyen müşterilerden oluşan müşteri gruplarının oluşturulması, her şeyden önce, kaynak tüketiminde verimliliği beraberinde getirmektedir. Bununla birlikte müşteri bölütlemesi, müşterilerin ihtiyaç ve davranışlarının tanınmasına olanak sağlayarak şirketler için yerinde çapraz satış (cross-sell) ve yukarı satış (up-sell) imkânları yaratmaktadır. Müşteri segmentasyonu gözetimsiz öğrenme (unsupervised learning) kategorisine dahil bir kümeleme (clustering) problemidir. Herhangi bir hedef değişkeni (target variable) bulunmayan bu problemler, bu özelliği ile gözetimli öğrenme (supervised learning) sınıfına dahil problemlerden ayrılmaktadır [1].

Müşteri kayıp analizi (customer churn analysis) mevcut müşterilerin kaybını önceden tahmin etmeye yönelik bir analiz biçimidir [2]. Bu analiz kullanılarak yapılan tahminlerle, müşteri kayıplarının önüne geçilmesi için müşteri ilişkileri yönetimi kapsamında çözümler üretilebilmesi mümkün olmaktadır. Müşteri kayıp analizi, şirketin aktif müşterilerinin şirket ile olan ilişkilerini sonlandırma olasılıklarının tahminlenmesine yönelik bir çalışmadır. Geçmişte çıkış yapmış olan müşteri veri setinin eğitici veri kümesini (training set) oluşturduğu müşteri kayıp analizi, gözetimli öğrenme (supervised learning) kategorisi altında incelenen bir problemdir. Sınırlı olan kaynaklar ile müşterileri tutundurmak için birçok kampanya yapılmaktadır. Araştırmamız kapsamında, müşteri kaçış olasılıkları hesaplanarak özel tutundurma kampanyaları yapılması

"978-1-7281-7206-4/20/$31.00 ©2020 IEEE"

hedeflenmektedir. Bununla beraber kaçış eğilimine girmiş her müşteri şirket için değil, sadece belli kanal ve üründeki müşterilerin kaçmasına yönelik çözümler geliştirilmesi de araştırmamız kapsamında irdelenmektedir.

III. ÖNERİLEN METODOLOJİ

Veri Ön İşleme Adımları: İş ihtiyaçlarına göre şekillenen müşteri segmentasyonu çözümü sürecimize, bireysel emeklilik sistemi (BES) müşterilerinin ayrıştırıcı niteliklerinin çözümlenmesiyle başlanmıştır. İçerisinde cinsiyet, medeni durum, meslek bilgisi, uyruk, fon dağılımı, giriş aidatı tipi gibi özniteliklerin bulunduğu 581 değişkenli bir veri seti türetilerek bu verinin ön analizi çalışmaları gerçekleştirilmiştir. Veri bilimi çalışmalarında karşılaşılan en büyük problemlerin başında temiz, güvenilir, kendi içerisinde tutarlı veriye erişim gelmektedir. Yapılan ön analiz çalışmaları neticesinde, gerek sistemsel (dizayn, entegrasyon vs.) gerekse de personele dayalı sebepler dolayısıyla tutarsız/kirli olduğu tespit edilen değişkenler analitik çalışmalar kapsamından çıkartılmıştır.

Kümeleme Analizi: Problem çözümünün ilk aşamasında, yukarıda bahsi geçen ön analizlerle beraber işi anlama çalışmaları neticesinde elde edilen 35 adet değişken ile esnek, farklı iş birimlerinin değişen ihtiyaçlarına doğrudan cevap verebilecek bir segmentasyon modeli geliştirilmiştir. Kullanılan veri setinde nümerik değişkenlerle birlikte cinsiyet, ayrılma talebi, ürün grupları gibi kategorik değişkenlerin bulunması ve uç değerlere (outlier) duyarsızlığı nedeniyle, segmentasyon yöntemi olarak "k-medoids" metodu benimsenmiştir [3]. "K-medoids" metodu, "k-means" yönteminden farklı olarak, ortalama değerlerin yerine veri setinde bulunan "medoid" adı verilen noktaları küme temsilcileri (cluster representatives) olarak kabul etmektedir [4]. Küme temsilcileri olarak veri setinde bulunan noktaların tercih edilmesi, metoda uç değerler konusunda gürbüzlük (robustness) kazandırmakla birlikte veri setinde nominal değerlerin kullanımını da mümkün kılmaktadır. Bununla birlikte bu tercih, metodu gerçekleyen algoritmaların veri setindeki tüm noktalar arası uzaklıkların önceden hesaplanarak bir uzaklık matrisi (distance matrix) bünyesinde tutulmasını gerektirmektedir.

Bu araştırma kapsamında veri setini kullandığımız firmanın 500.000'den fazla aktif BES müşterisinin bölümleneceği dikkate alındığında, yukarıda bahsi geçen uzaklık matrisinin boyutu 1000 GB'ı kolaylıkla aşmaktadır. Bu araştırma kapsamında önerdiğimiz çözüm yaklaşımında; böylesi bir bellek ihtiyacı problemi, rastgele örnekleme (random sampling) yöntemini temel alarak büyük veri setleri üzerinde k-medoids metodunu uygulanabilir kılan CLARA (Clustering Large Applications) çözümü vasıtasıyla aşılmaktadır [4]. Araştırmamızın ilerleyen adımlarında saha çalışmaları kapsamında firma ile yapmış olduğumuz ihtiyaç/talep toplantıları neticesinde, sektörün temel gereksiniminin "müşteri değer segmentasyonu" olduğu netleştirilmiştir. Bu çerçevede kurum dahilinde ön analiz aşamasında yapılan anket çalışmaları analiz edilmiş ve müşteri değer tanımları ile bu tanımlara etki ettiği düşünülen değişkenler ortaya konulmuştur. Müşteri değerini belirleyen faktörler sırasıyla "katılımcı katkısıyla oluşan toplam fon büyüklüğü", "son tahsilatlardan oluşturulmuş, aylığa indirgenmiş katkı payı", "düzenli ödenen ürünlere ait tahsilat oranı", "müşteri olma süresi", "müşteriden elde edilen kesintiler toplamı" olarak belirlenmiştir.

Yapmış olduğumuz k-medoids [5] ile müşteri segmentasyonu çalışmasına ek olarak farklı iki kümeleme metodu üzerinde durulmuştur: k-means [6] ve DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [7]. Olasılıksal olmayan bir kümeleme metodu olan k-means, genel itibariyle NP-hard bir probleme çözüm arıyor olsa da verimli sezgisel yaklaşımlar vasıtasıyla hızlı bir biçimde lokal optimumlara yakınsar ve bu özelliği dolayısıyla en çok tercih edilen kümeleme metotlarından biri olmuştur ([8], [9]). Bununla birlikte iteratif yollardan çözüm bulan k-means yöntemi, her bir iterasyon sonucunda, ilgili kümelerin tanımlayıcısı/merkezi olarak o aşamada o kümeye ait noktaların ortalama değerini kabul eder. Bu durum metodun, verideki uç değerlere olan hassasiyetini artırmaktadır. Bu hassasiyeti dikkate alarak çalışmamıza veri setinde yer alan muhtemel uç değerleri eleyerek başlanmıştır. Tek değişkenli (univariate) analiz sonucunda, her bir değişken için uç değer sayılabilecek girdiler belirlenmiş, bunları ihtiva eden müşteriler analizden çıkabilmiştir.

Çalışmamızda kullanmış olduğumuz k-means ve k-medoids metotları, temel olarak, küresel (spherical) ve eşit büyüklükte kümelerin ayrıştırılmasında başarılı sonuçlar vermektedir. Veri setimizde bulunması muhtemel aksi yapıların tespit edilebilmesi maksadıyla, özellikle uzamsal uygulamalarda oldukça başarılı sonuçlar üreten, DBSCAN metodu da çalışmamıza dahil edilmiştir. DBSCAN parametrik olmayan, yoğunluk temelli bir kümeleme algoritması olup uç değerleri algoritmanın dizaynı vesilesiyle içsel olarak değerlendirir ve bu değerlere karsı gürbüz bir metottur. Kompleks yapıların birbirinden ayrıştırılabilmesine olanak sağlayan DBSCAN, kullanıcının belirlediği "minPts" ve "eps" isimli parametreler uyarınca çözüme gider. Veri setindeki toplam küme sayısı ve uç değerler, kullanıcı tarafından girilen bu iki parametre tarafından (dolaylı olarak) belirlenir [7]. Oluşturulan segmentler sonrasında segmentler arası göç analizleri ve alarmları geliştirilerek düzenli periyotta (aylık) çalıştırılan segment modellerinin çıktıları tasarlanan segment Data Mart'da tarihsel olarak saklanmaktadır. Bu çıktılar arasında; segment bilgisi, müşteri değişkenlerinin bilgisi ve değerleri, segmentlerin betimsel istatistikleri (Min, Max, Median, Std.Dev.). Ayrıca data mart üzerinden çeşitli analizler ve alarmlar SQL yardımıyla türetilmektedir. Bunlar, müşteri segment göçlerinin analizi; bir önceki çalışan segment çıktıları ile müşteri bazlı karşılaştırma yapıp segmenti değişen müşteriler için alarm türetilmesi. Müşteri bazlı segment haritası; tarihsel olarak tüm segmentlerin analiz edilmesi, segment modellerinin betimsel istatistiklerinin analiz edilmesi.

Çalışmamızın müşteri bölütleme bölümünde olduğu gibi, müşteri kayıp analizi çalışması da müşteri kaybını müşterinin sadakatinden ayrıştıracak özniteliklerin tespiti ve bu değişkenlerin ön analizi ile başlamıştır. Bu kapsamda "Müşteri – Ürün Evrensel" Datamart'ındaki değişkenler analiz edilmiştir. Bu çalışmalar neticesinde şu değişkenler kurulacak olan modellere girdi olarak kabul edilmiştir: Güncel aylığa indirgenmiş katkı payı, Tahsil edilmiş olan giriş aidatı, Müşterinin aktarımla gelip gelmediği bilgisi, Müşteri yaşam süresi, Yönetim gider kesintisi, Müşterinin BES'e dahil edildiği acente, Güncel devlet katkısı, Cinsiyet, Doğum yeri, Yaşadığı il,

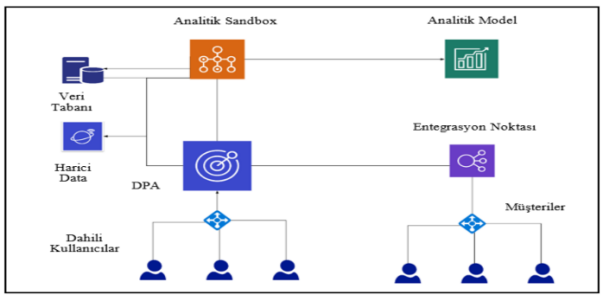
Doğum tarihi. Ulaşabildiğimiz, modellemede kullanılması olası veri setini ele aldığımızda geçmişte çıkış yapmış müşterilerin, mevcut yaşayan müşterilere oranının düşük oluşu, elimizdeki problemin dengesiz (imbalanced) bir tahminleme problemi olduğunu göstermektedir [10]. Buna ek olarak, ilerleyen zaman içerisinde değişen sektörel dinamiklerin de göz önüne alınması ihtiyacı, çalışmamızda kullanılacak eğitici ve test veri kümelerinin eşit zaman dilimlerine bölünerek modellemelerde değerlendirilmesi ihtiyacını doğurmuştur. Daha net bir ifadeyle, analize konu olan tüm müşteriler (yaşayıp yaşamadığından bağımsız olacak şekilde), sisteme giriş tarihlerine göre kesikli (discrete) hale getirilmiş zaman grupları (time interval) içerisine dahil edilmiştir. Her bir tarih aralığı için ayrı bir model, yine o tarih aralığına dahil olan müşteri verileri üzerinden eğitilmiş, modellerin değerlendirmeleri de yine o tarih aralığına ait müşteriler üzerinden yapılmıştır. Eğitici veri kümeleri oluşturulurken, ilgili zaman aralığındaki çıkış yapmış müşteri adetleri baz alınmış, karşılık olarak aynı gruba dahil aynı sayıda negatif (çıkış yapmamış) örnek eğitici kümeye rastgele seçimle dahil edilmiştir. Dış etkenlerin etkilerini modele dahil etmek maksadıyla döviz kur artışı oranı dahil edilmiştir. Ay bazında elde edilen faiz bilgiler müşterilerin, başlangıç ve bitiş tarihlerini göz önünde bulundurularak yaşadıkları dönem aralığındaki döviz artış miktarı oranı hesaplanmış ve yeni bir değişken olarak eğitici ve test veri kümelerine eklenmiştir. Ayrıca korelasyon matrisinde diğer değişkenler ile yüksek pozitif veya negatif ilişkisi olmadığı gözlemlenmiştir.

Müşteri kayıp analizi çalışmalarında sırasıyla lojistik regresyon (logistic regression) [11], destek vektör makineleri (support vector machines, SVM) [12], yapay sinir ağları (artificial neural networks, ANN) [13] ve XGBoost (extreme gradient boosting) [14] yöntemleri kullanılmıştır. Olasılıksal fark gözeten (probabilistic discriminative model) öğrenme modellerinden biri olan lojistik regresyon, basit matematiksel yapısı ve sonuçlarının anlamlandırmaya (model interpretation) oldukça müsait yapısı nedeniyle, özellikle olasılık tahminlemeyi gerektiren gözetimli öğrenme problemlerinde en çok başvurulan modellerin başında gelmektedir [15]. Olasılıksal olmayan ikili lineer sınıflandırıcı (nonprobabilistic binary linear classifier) olarak lineer destek vektör makinelerini kullanarak elde ettiğimiz müşteri kayıp modelinin çıktıları aşağıdaki tabloda verilmiştir. Bu modellere ek olarak müşteri kayıp analizi çalışmamızda beslemeli yapay sinir ağları kategorisinde değerlendirilen çok katmanlı algılayıcı model (multilayer perceptron) değerlendirilmiştir. Giriş (input), gizli (hidden) ve çıkış (output) katmanları (layer) olmak üzere üç kademeden oluşan bu yapay sinir ağı modeli, gözetimli öğrenme biçimlerinden geri yayılımı (backpropagation) kullanarak kendisini optimize etmektedir. Son yıllarda, veri bilimi proje yarışmalarında en iyi sonuçları alması ile popülerleşen XGBoost, gradyan hızlandırılmış karar ağaçları (gradient boosted decision trees) konseptinin bir uygulaması (implementation) olarak karşımıza çıkmıştır. XGBoost temelde zayıf öğrenicileri (weak learner) bir araya getirerek (ensemble) karar ağaçları oluşturur [16]. Metodolojinin gerçeklenmesi aşamasında farklı teknoloji ve altyapılar kullanılarak sistem mimarisi oluşturulmuştur. Kullanıcı arayüzünün oluşturulmasında; Java, J2EE, Spring Boot, HTML5, Jquery, JPA Hibernate, PostgreSQL teknolojileri kullanılmış, ayrıca veri analizi için oluşturulan analytic sandbox'da R, Apache Pig Data, Livy, Spark, Hadoop, Docker ortamları kullanılmıştır.

Veri ve veriyi anlatan üstverinin yönetiminin nasıl sağlandığı üzerinde farklı araştırmaların literatürde yer aldığı görülmektedir [18-20]. Dağıtık veri saklama platformları için bilgi sistemleri alanında ve dağıtık ortamda verinin aranması alanında da araştırmalar yapıldığı görülmektedir [21-26]. Bu araştırmalardan farklı olarak, bu çalışma kapsamında, veriden müşteri kayıp analizi tahmini amaçlı anlam çıkartılması üzerinde odaklanılmıştır.

IV. Önerilen Metodolojinin Prototip Uygulaması

Müşteri segmentasyonu ve müşteri kayıp eğilimi analizlerinde şirket tarafından sağlanan veri setleri kullanılmıştır. K-means metodunda, bölüt sayısı kullanıcı tarafından belirlenmesi gerekmektedir [26]. Saha çalışmaları kapsamında yapmış olduğumuz çalışmalar ve toplantılar neticesinde toplam müşteri kümesi adetinin 4 olması kararlaştırılmıştır. (Sektörün segmentler bazında hizmette farklılaşma kabiliyeti bakışıyla değerlendirmeler yapılmıştır).



Şekil. 1. Sistem mimarisi

V. Önerilen Metodolojinin Değerlendirilmesi

Bununla birlikte, veri üzerinde yapmış olduğumuz analiz neticesinde (aşağıdaki toplam "within sum of squares" değerinin bölüt sayısına göre grafiği incelenerek) optimum bölüt sayesinin 4 ile 8 arasında olabileceği tespit edilmiştir. Her iki durumu da göz önüne alarak, çalışmamızda müşteri bölüt sayısının 5 olmasına karar verilmiştir. K-means yöntemine ek olarak, k-medoids kümeleme metodu da uygulanmıştır. CLARA algoritması ile gerçeklenen bu yöntemde, her bir liberasyonda (50 adet) 500.000 civarı müşteri içerisinden 10.000'lik rastgele setler seçilmiş, bu setlerden elde edilen (optimum) melodilere göre tüm veri seti kümelenmiştir. DBSCAN, müşteri olma süreleri bakımından ayrıştırılamayan müşterileri tahsilat oranı, aylık katkı payları ve toplam fon büyüklükleri açısından yüksek değerli ve düşük değerli olmak üzeri iki ana kategoriye ayırdığı gözlemlenmiştir. Müşteri kayıp eğilimi analizinde kullanılan algoritmaların hata matrisi (Confusion matrix) ve değerlendirme parametre sonuçları aşağıdaki tablolarda belirtilmiştir.

TABLO I. HATA MATRİSİ

Algoritma	Genel Skor	Hata Matrisi	Hayır Oranı	Evet Oranı	R^2 Skoru	ROC-AUC Skoru
Lojistik Regresyon	0.808	[10834 2787] [2423 11123]	%80.0	%82.0	0.232	0.808
Destek Vektör Makineleri	0.810	[10662 2959] [2177 11369]	%78.0	%84.0	0.243	0.811
Yapay Sinir Ağları	0.829	[11729 1892] [2729 10817]	%86.0	%80.0	0.319	0.829
XGBoost	0.851	[12009 1612] [2412 11134]	%88.0	%82.0	0.407	0.851

VI. SONUÇLAR VE GELECEKTEKİ ÇALIŞMALAR

Çalışmamızda müşteri bölütlemesinde sıklıkla kullanılan gözetimsiz öğrenme algoritmaları k-medoid, k-means ve DBSCAN algoritmaları da kullanılmış, sonuçlar iş bilgisi ile değerlendirilerek anlamlandırılmıştır. Ayrıca segmentler arası geçiş analizleri de yazılıma dahil edilerek müşteri yaşam süreci gözlemleme imkânı sağlanmıştır. Müşteri kayıp analizinde ise iş bilgisi kullanılarak üretilen değişkenler üzerinden sınıflandırma algoritmaları koşularak 0.80'in üzerinde başarı oranları elde edilmiştir. Çalışma sayesinde segmentler arasında karşılaştırılabilir müşteri kayıp oranları oluşturularak kişiselleştirilebilir tutundurma faaliyetleri oluşturabilme imkânı elde edilmiştir. Çalışmanın sonraki aşamalarında, sınıflandırma algoritmalarının başarı oranlarını artırmaya yönelik çalışmaların yapılması planlanmaktadır. Ayrıca segmentler arası geçiş süreci, müşteri kayıp analizinde bir değişken olarak kullanılabileceği değerlendirilmektedir.

TEŞEKKÜR

Bu çalışma 7180575 proje numarası ile Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından desteklenmiştir. Çalışmamızı finansal olarak destekleyen TÜBİTAK'a teşekkürlerimizi sunarız. Ayrıca gerekli çalışma ortamı ve teknolojik desteği sağlayan DİP Danışmanlık'a, Dr. Mehmet Aktaş'a ve DİP Danışmanlık personeli Orhan Toprakman'a teşekkür ederiz.

KAYNAKLAR

[1] Han, J., Kamber, M. & Pei, J. (2012). Data mining concepts and techniques, third edition Morgan Kaufmann Publishers

[2] Eria, Kamya & Poolan Marikannan, Booma. (2018). Systematic Review of Customer Churn Prediction in the Telecom Sector. 2. 7-14.

[3] Tripathi, Shreya & Bhardwaj, Aditya & E, Poovammal. (2018). Approaches to Clustering in Customer Segmentation. International Journal of Engineering & Technology. 7. 802. 10.14419/ijet.v7i3.12.16505.

[4] Kaufman, Leonard & Rousseeuw, Peter. (2009). Finding Groups in Data: An Introduction to Cluster Analysis.

[5] Aryuni, Mediana & Madyatmadja, Evaristus & Miranda, Eka. (2018). Customer Segmentation in XYZ Bank Using K-Means and K-Medoids Clustering. 1-9. 10.1109/ICIMTech.2018.8528086.

[6] Mihova, Vesela & Pavlov, Velisar. (2018). A customer segmentation approach in commercial banks. AIP Conference Proceedings. 2025. 030003. 10.1063/1.5064881.

[7] Ester, Martin, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." KDD (1996).

[8] S. Lloyd, "Least squares quantization in PCM," in IEEE Transactions on Information Theory, vol. 28, no. 2, pp. 129-137, March 1982.DOI: 10.1109/TIT.1982.1056489

[9] Bishop, Christopher M. Pattern Recognition and Machine Learning. New York: Springer, 2006.

[10] Ali, Özden & Arıtürk, Umut. (2014). Dynamic churn prediction framework with more effective use of rare event data: The case of private banking. Expert Systems with Applications. 41. 7889–7903. 10.1016/j.eswa.2014.06.018.

[11] Lemmens, Aurélie & Gupta, Sunil. (2019). Managing Churn to Maximize Profits". SSRN Electronic Journal. 10.2139/ssrn.2964906.

[12] Hossain, Md Mosharaf & Miah, Mohammad. (2015). Evaluation of different SVM kernels for predicting customer churn. 1-4. 10.1109/ICCITechn.2015.7488032.

[13] Khan, Yasser & Shafiq, Shahryar & Naeem, Abid & Ahmed, Sheeraz & Safwan, Nadeem & Hussain, Sabir. (2019). Customers Churn Prediction using Artificial Neural Networks (ANN) in Telecom Industry. International Journal of Advanced Computer Science and Applications. 10. 10.14569/IJACSA.2019.0100918.

[14] Ahmad, A.K., Jafar, A. & Aljoumaa, K. J Big Data (2019) 6: 28. https://doi.org/10.1186/s40537-019-0191-6

[15] Duda, Richard & Hart, Peter & G.Stork, David. (2001). Pattern Classification.

[16] Chen, Tianqi & Guestrin, Carlos. (2016). XGBoost: A Scalable Tree Boosting System. 785-794. 10.1145/2939672.2939785.

[17] Baeth, M.J., Aktas, M.S., (2018) . An approach to custom privacy violation detection problems using big social provenance data, Concurrency and Computation-Practice & Experience, Vol: 30, Issue:21.

[18] Baeth, M.J., Aktas, M.S., (2019). Detecting misinformation in social networks using provenance data, Concurrency and Computation-Practice & Experience, Vol: 31, Issue:3.

[19] Riveni, M. et al. (2019). Application of provenance in social computing: A case study, Concurrency and Computation-Practice & Experience, Vol: 31, Issue:3.

[20] Tas, Y., Baeth M.J., Aktas, M.S., (2016) An Approach to Standalone Provenance Systems for Big Provenance Data, The International Conference on Semantics, Knowledge and Grids on Big Data (SKG-16).

[21] Aktas, M.S., (2018) Hybrid cloud computing monitoring software architecture, Concurrency and Computation: Practice and Experience, Vol: 30, Issue:21.

[22] Aktas, M.S. et al. (2004). A web based conversational case-based recommender system for ontology aided metadata discovery, The 5th IEEE/ACM International Workshop on Grid Computing, pp:69-75.

[23] Aktas, M.S. et al, (2007). Fault tolerant high-performance Information Services for dynamic collections of Grid and Web services, Future Generation Computer Systems, Vol:23, Issue: 3.

[24] Pierce, M.E. et al. (2008). The QuakeSim project: Web services for managing geophysical data and applications, Pure and Applied Geophysics, Vol:165, Issue: 3-4, pp. 635-651.

[25] Aydin, G. et al. (2005). SERVOGrid complexity computational environments (CCE) integrated performance analysis, The 6th IEEE/ACM International Workshop on Grid Computing.

[26] Jin, Xin & Han, Jiawei. (2017). K-Means Clustering. 10.1007/978-1-4899-7687-1_431.