

Müşteri Segmentasyon Sınıflandırması

Customer Segmentation Classification

Alperen USLU
Bilgisayar Mühendisliği
Kahramanmaraş Sütçü İmam Üniversitesi
Kahramanmaraş, Türkiye
alperennuslu@gmail.com

Özetçe —Bir otomobil şirketi, mevcut ürünleri ile yeni pazarlara girmeyi planlıyor. Mevcut pazarlarında, satış ekibi tüm müşterileri 4 segmente ayırdı ve bu strateji onlar için olağanüstü iyi çalıştı. Şimdi, aynı stratejiyi yeni pazarlarda kullanmayı doğru gruba tahmin etmeyi planlıyorlar. Bu, müşteri segmentasyonu uygulamasıdır, yani bir müşteri tabanını yaş, cinsiyet, ilgi alanları ve harcama alışkanlıkları gibi belirli yönlerden benzer olan birey gruplarına bölmek. Bu çalışmada bahsi geçen problemin çözümü için bir çok makine öğrenmesi modelleri ve öznelik seçim yaklaşımı önerilmiştir. Önerilen yöntem, literatüre yaygın kabul gören destek vektör makinesi, K en yakın komşu, Naive Bayes, karar ağaçları, lojistik regresyon, random forest, xgboost, lightgbm, catboost ve gbm sınıflandırma yöntemleriyle karşılaştırılmalı olarak analiz edilmiştir.

Anahtar Kelimeler—Segment, müşteri

Abstract—A car company is planning to enter new markets with its existing products. In their current markets, the sales team has divided all customers into 4 segments, and this strategy has worked extraordinarily well for them. Now, they plan to use the same strategy in new markets to predict the correct group. This is a customer segmentation application, which is to divide a customer base into groups of individuals that are similar in specific ways such as age, gender, interests, and spending habits. In this study, many machine learning models and feature selection approaches have been proposed to solve the problem mentioned. The proposed method has been analyzed comparatively with the support vector machine, K nearest neighbor, Naive Bayes, decision trees, logistic regression, random forest, xgboost, lightgbm, catboost, and gbm classification methods, which are widely accepted in the literature.

Keywords—Segmentation, customer

I. GİRİŞ

Müşteri Segmentasyonu, bir müşteri tabanını, paylaştıkları belirli özelliklere göre kişi gruplarına ayırmayı içerir.

Segmentasyon, bir şirketin müşteri ilişkileri yönetimine ve genel pazarlama performansına yardımcı olur. Müşteri tabanınızı bölümlere ayırarak, şirketlerin pazarlama çabalarını ilgili mesajlarla belirli grupları hedef alacak şekilde uyarlamasına olanak tanır. Bu, şirketin pazarlama kaynaklarını daha iyi tahsis etmesine, etkili pazarlama kampanyası stratejisi oluşturmaya ve daha yüksek dönüşümle istenen pazarlama sonuçlarını en üst düzeye çıkarmasına olanak tanır [?].

Müşteri segmentasyonuna yönelik ek bir yaklaşım, yeni segmentleri keşfetmek için makine öğrenimi algoritmalarından

yararlanmaktır. Pazarlamacıların tasarladığı segmentasyon modellerinden farklı olarak, makine öğrenimi müşteri segmentasyonu, gelişmiş algoritmaların, pazarlamacıların kendi başlarına keşfetmekte zorluk çekebilecekleri içgörülerini ve gruplamaları ortaya çıkarmasına olanak tanır [?].

Ayrıca, segmentasyon modeli ile kampanya sonuçları arasında bir geri bildirim döngüsü oluşturan pazarlamacılar, müşteri segmentlerini sürekli geliştirecek. Bu durumlarda, makine öğrenimi modeli yalnızca segment tanımını iyileştirmekle kalmayacak, aynı zamanda segmentin belirli bir alt kümesinin diğerlerinden daha iyi performans gösterip göstermediğini belirleyerek pazarlama performansını optimize edebilecektir [?].

Bu tür bir segmentasyon, kuruluşların pazarlama çabalarını hedef kitlenin çeşitli alt gruplarına göre uyarlamalarına olanak tanır. Bu çabalar hem iletişim hem de ürün geliştirme ile ilgili olabilir.

Ayrıca, müşteri segmentasyonunun aşağıdaki faydaları vardır:

Kimi hedeflediğinizi bildiğinizde, belirli müşteri gruplarını etkileyecek hedefli pazarlama mesajları oluşturabilirsiniz. Müşteri segmentasyonu, e-posta, sosyal medya gönderileri, web sitesi reklamları veya diğer geleneksel medya yoluyla segment için en iyi iletişim kanalını seçmenize olanak tanır. Ürün ve hizmetleri iyileştirmenin yollarını belirleyin. Kuruluşa en çok para yatıran müşterilerin gerçek ihtiyaçlarına dayanır. Yaşları, konumları, satın alma alışkanlıkları, ilgi alanları vb. hakkında bilgi sahibi olarak çabaları en karlı müşterilere yönlendirin. Bu şekilde, doğru kitleyi çekmek için doğrudan pazarlama mesajları göndermek mümkündür. Daha iyi müşteri ilişkileri kurun ve marka sadakati oluşturacak şekilde bağlantı kurun. Kurumlar, işlerinin en önemli segmentinin ihtiyaçlarını karşıladıktan sonra kişiselleştirilmiş müşteri hizmetleri sunabilir ve böylece markalarını sektörde konumlandırabilirler.

Bu çalışmada, bir otomobil şirketi mevcut ürünleri ile yeni pazarlara girmeyi planlamaktadır. Yoğun pazar araştırması sonucunda, yeni pazarın davranışının mevcut pazarlarına benzer olduğunu belirlemişlerdir.

Mevcut pazarlarında, satış ekibi tüm müşterileri 4 segmente (A, B, C, D) ayırmıştır. Daha sonra, müşteri segmentlerine yönelik segmente edilmiş iletişim ve ulaşım gerçekleştirdiler. Bu strateji onlar için olağanüstü iyi çalışmıştır. Aynı stratejiyi

yeni pazarlar için kullanmayı planlıyorlar ve 2627 yeni potansiyel müşteri belirlediler.

Yeni müşterilerin doğru grubunu tahmin etmesinde yardımcı olmanız gerekmektedir. Bu süreçte, 10 farklı model kullanmayı planlıyorum. Bu modeller, veri setindeki performanslarını değerlendirmek için kullanılacaktır. Bu modellerin seçimi ve uygulanması, müşteri segmentasyon stratejisinin başarısında kritik bir rol oynayacaktır. Bu nedenle, bu çalışmanın amacı, en etkili modelin seçilmesine yardımcı olmaktır.

II. YÖNTEM

Müşteri segmentasyonu işlenmesi sürecinde, müşteri verileri ham veri olarak kullanılır. Bu veriler, müşterinin satın alma geçmişi ve bilgileri olabilir. Ham verilerden gürültülerin filtrelene süreci ön işlem olarak adlandırılır. Bu aşamada, eksik veriler doldurulabilir, aykırı değerler düzeltilir ve veriler uygun bir formata dönüştürülebilir. Filtrelenmiş veriden çeşitli öznitelik çıkarma algoritmalarıyla, problemi temsil edebilecek veri oluşturulur. Elde edilen bu veriler bir sınıflandırma algoritması ile müşteri segmentasyonu yapılır. Ancak, bu genel yaklaşımda, öznitelik çıkarma algoritmalarından elde edilen verideki bazı vektörlerin probleme özgü olarak anlamlı ya da anlamsız olduğuna bakılmaksızın kullanılır. Bu durum müşteri segmentasyonunun başarısını oldukça etkileyebilir. Bu nedenle, öznitelik seçimi sürecinde, probleme özgü anlamlı özniteliklerin seçilmesi önemlidir. Bu, modelin performansını artırabilir ve daha doğru müşteri segmentasyonu sağlayabilir.

A. Öznitelik Seçimi

Öznitelik seçimi sürecinde, veri ön işleme adımlarında bir dizi teknik kullandım: Label Encoder: Kategorik değişkenleri sayısal değerlere dönüştürmek için Label Encoder kullandım. Bu, makine öğrenmesi modellerinin kategorik verileri daha iyi anlamasını sağlar. Her bir benzersiz kategori değeri, benzersiz bir tam sayıya eşlenir.

StandardScaler: Veri setindeki özellikleri standardize etmek için StandardScaler kullandım. Bu, tüm özelliklerin aynı ölçekte olmasını sağlar, böylece hiçbir özellik diğerlerinden daha baskın hale gelmez. StandardScaler, her özelliği, ortalama 0 ve standart sapma 1 olacak şekilde yeniden ölçeklendirir.

Get Dummies: Kategorik değişkenleri “dummy” veya “indicator” değişkenlere dönüştürmek için kullandım. Bu, her bir kategori değerini yeni bir özellik olarak temsil eder ve bu özellik, belirli bir kategori değerinin varlığını veya yokluğunu gösterir.

Bu teknikler, modelin verileri daha etkili bir şekilde anlamasına ve daha doğru tahminler yapmasına yardımcı olur. Her bir teknik, veri setinin belirli bir yönünü ele alır ve genel model performansını iyileştirir. Bu adımların uygulanması, modelin genel performansını ve doğruluğunu artırırken, aynı zamanda modelin eğitim süresini de optimize eder. Bu nedenle, bu tekniklerin kullanılması, veri bilimi projelerinde önemli bir adımdır.

III. DENEYSEL SONUÇLAR VE TARTIŞMA

Deneysel sonuçları elde edebilmek için kaggle sitesinde 2023 yılında paylaşılmış olan "Customer Segmentation Classification" veri seti [?] kullanılmıştır. Elde edilen sonuçlar, google colab [?] üzerinden elde edilmiştir.

Bu bölümde, öncelikle kullanılan veri setinden bahsedilmekte, daha sonra karşılaştırmalı olarak elde edilen simülasyon sonuçları sunulmaktadır.

A. Müşteri Segmentasyon Sınıflandırma veri seti

Otomobil şirketi, mevcut ürünleri yeni pazarlara girmeyi planlıyor. Yoğun pazar araştırması sonucunda, yeni pazarın davranışının mevcut pazarlarına benzer olduğunu belirlemişler. Mevcut pazarlarında, satış ekibi tüm müşterileri 4 segmente ayırmıştır. Daha sonra, müşterilerin farklı segmentleri için segmente edilmiş erişim ve iletişim gerçekleştirdiler. Bu veri seti 2020 yılında bir hackathon [?] deposundan paylaşılmıştır. Müşteri segmentasyonu uygulayan şirketler, her müşterinin farklı olduğu ve tüketicilerin alakalı bulacağı ve onları bir şey satın almaya yönlendireceği mesajlarla belirli, daha küçük grupları hedeflerlerse pazarlama çabalarının daha iyi sonuç vereceği gerçeğiyle çalışırlar. Şirketler ayrıca, pazarlama materyallerini o segmente daha doğru şekilde uyarlamak için her segmentin neyi en değerli bulduğunu keşfetme fikriyle müşterilerinin tercihleri ve ihtiyaçları hakkında daha derin bir anlayış kazanmayı umuyor. Veriseti için 18 yaşından büyük 10695 kişi katılmıştır.

B. Simülasyon Sonuçları

Veri işleme adımlarında sonra oluşan veriler ile sınıflandıma modelleri için, veri seti üzerinden SVM, KNN, DT, NB, LR, RF, XGB, LGBM, CatBoost ve GBM sınıflandırıcıları kullanılmıştır. Gerçekleştirilen bu modellerden, *acc*, *AUC*, *recall*, *Precision*, *F1* değerleri kaydedilmiştir. Bu değerlerin sonuçları Tablo I, Tablo II’te verilmiştir.

	Acc	Auc	recall
Logistic Regression (LR)	47.78	72.35	46.84
K-Nearest Neighbors (KNN)	41.47	65.59	40.93
Decision Tree (DT)	39.93	60.02	39.50
Naive Bayes (NB)	44.65	70.80	44.62
Support Vector Machine (SVM)	48.80	72.97	48.13
Random Forest (RF)	44.41	70.33	43.80
XGBoost	47.36	73.43	46.89
LightGBM	47.83	74.48	47.31
CatBoost	48.34	74.21	47.84
Gradient Boosting (GB)	49.98	75.24	49.37

Tablo I: Makine öğrenmesi modelleri *acc*, *AUC* ve *recall* ölçütlerinin sonuçları

Tablo I incelendiğinde; 49,98 ile en yüksek ortalama doğruluk değerine Gradient Boosting algoritması sahipken, daha sonra SVM algoritması ile ortalama 48,80 doğruluk değeri elde edilmiştir. Catboost ve LR yöntemleri sırasıyla 48,34; 47,78 ortalama doğruluk değerleri elde edilmiştir. Doğruluk değerleri incelendiğinde, diğer yöntemler ile Gradient Boosting arasında, Gradient Boosting algoritmasının lehine önemli bir fark olduğu görülmektedir. Aynı zamanda, GB algoritması diğer yöntemlere göre oldukça üstün bir doğruluk değeri elde ettiği Tablo I’de görülmektedir.

	Precision	F1
Logistic Regression (LR)	47.08	44.69
K-Nearest Neighbors (KNN)	41.92	41.24
Decision Tree (DT)	39.62	39.56
Support Vector Machine (SVM)	48.09	47.72
Random Forest (RF)	43.56	43.63
XGBoost	46.55	46.59
LightGBM	47.13	47.02
CatBoost	47.48	47.54
Naive Bayes (NB)	43.34	41.63
Gradient Boosting (GB)	49.14	48.87

Tablo II: Makine öğrenmesi modelleri *precision* ve *F1* ölçütlerinin sonuçları

Sınıflandırma algoritmalarının başarılarını karşılaştırmada en önemli ölçütlerden olan *AUC* değerlerinde de GB algoritmasının, diğer yöntemlere göre üstün bir başarı elde ettiği görülmektedir. *precision* ve *F1* açısından Tablo II incelendiğinde, en yüksek F1 skoru Gradient Boosting modeline ait. Bu model, diğer tüm modellerden daha iyi bir harmonik ortalama sağlar, yani hem Precision hem de Recall değerlerini dengeli bir şekilde optimize eder. En yüksek Precision değeri yine Gradient Boosting modelinde görülüyor. Bu, modelin pozitif olarak tahmin ettiği durumların ne kadarının gerçekten pozitif olduğunu gösterir.

Karşılaştırma ölçütleri genel olarak değerlendirildiğinde, Gradient Boosting algoritması rakip algoritmalara göre üstünlük sağladığı anlaşılmaktadır.

IV. SONUÇ

Bu çalışmanın sonuçları, otomobil şirketi için yeni pazarlara genişleme stratejisinin etkinliğini göstermektedir. Ancak, bu stratejinin başarısını daha da artırmak için, şirketin müşteri segmentasyon modelini sürekli olarak güncellemesi ve iyileştirmesi gerekmektedir. Bu, şirketin pazar dinamiklerine hızlı bir şekilde uyum sağlamasını ve müşteri ihtiyaçlarını daha iyi karşılamasını sağlar. Bu nedenle, bu alandaki gelecek araştırmalar, şirketin müşteri tabanını genişletme ve satışlarını artırma yeteneğini daha da geliştirebilir. Bu, hem şirketin karlılığı hem de müşteri memnuniyeti açısından önemlidir.