



AKILLI BAKIMDA YENİ BİR YAKLAŞIM: GERÇEK ZAMANLI YEDEK PARÇA TAHMİNLEME

Eren Esgin^{1,2} , Volkan Özay¹, Görkem Özkan¹

¹ Yapay Zeka Araştırmaları, MBIS ArGe Merkezi, İstanbul, Türkiye

² Enformatik Enstitüsü, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

{eren.esgin, volkan.ozay, gorkem.ozkan}@mbis.com.tr, eren.esgin@metu.edu.tr

ÖZET

Satış sonrası müşteri hizmetleri, kaliteli yedek parça hizmetleri ve uygun fiyatlar ile hedef müşteri kitlesini memnun etmeye çalışmaktadır. Sıkça tekrarlayan müşteri ziyaretleri operasyonel satış maliyetlerinde önemli artışlara neden olurken, bu yanıltıcı politika uzun vadede kurumsal itibarı da zedeleyebilir. Akıllı bakım bağlamında gerçek zamanlı yedek parça tahminleme çözümü, C4.5, Apriori algoritmaları ve ağırlıklandırılmış k-en yakın komşuluk (kNN) uyarlamalarını birleştiren hibrit bir sınıflandırma modeli ile sorumlu teknik servise ilgili arıza için en olası yedek parçayı proaktif olarak önerir. Deneysel sonuçlara göre, önerilen yaklaşımın insan düzeyindeki yedek parça tahminleme performansını yaklaşık iki katına çıkardığı görülmektedir. En iyi model konfigürasyonuna göre televizyon ürün grubu için gerçek zamanlı yedek parça tahminleme çözümü SAP sisteminde canlı kullanıma alınmıştır.

Anahtar Kelimeler: Sınıflandırma, Akıllı Bakım, SAP, Gerçek Zamanlı Yedek Parça Tahmini, Ağırlıklandırılmış k-En Yakın Komşuluk.

A NEW APPROACH IN INTELLIGENT MAINTENANCE: SPARE PART PREDICTION ON REAL TIME BASIS

ABSTRACT

After sales customer services attempt to satisfy target customers by qualified spare part services and affordable prices. While undesirable repetitive customer visits result in significant increase at sales operational costs, this misleading policy may deteriorate organizational goodwill at long run. In the context of intelligent maintenance, spare part prediction solution proactively proposes the most probable spare part to the responsible technical service by a hybrid classification model that combines C4.5, Apriori algorithms and weighted k-Nearest Neighbor (kNN) adaptations. The experimental results demonstrate that proposed approach approximately doubles the human-level performance at spare part prediction scenario. According to best runtime configuration analysis, a real time spare part prediction model has been deployed for television product group at the client's SAP system.

Keywords: Classification, Intelligent Maintenance, SAP, Spare Part Prediction on Real Time Basis, Weighted k-Nearest Neighbor.

1. GİRİŞ

(Domingos, 2015)'de verilen araştırma sonuçlarına göre işletmeler ortalama brüt gelirlerinin %24'ünü satış sonrası müşteri hizmetlerinden elde ederken, brüt kârın %45'ini bu iş kolundan sağlamaktadır. Bu gerçeğe paralel olarak, arıza başına ortalama müşteri ziyareti sayısı, satış sonrası müşteri hizmetlerinde kritik bir anahtar performans göstergesidir. Şöyle ki, sıkça tekrarlanan çözümsüz müşteri ziyaretleri, operasyonel satış giderlerinde (örn. garanti, bakım ve nakliye maliyeti) ciddi bir artışa neden olur. Ayrıca bu istenmeyen müşteri ziyaretleri, satış sonrası müşteri hizmetlerinin kalite düzeyini etkileyebilmekte ve uzun vadede kurumsal marka değerine zarar verebilmektedir. Yedek parça tahminleme iş senaryosu, arıza özelliklerine, ürün özelliklerine, müşteri profiline ve teknik servis özelliklerine göre yedek parça tüketim örüntülerini somutlaştırıp, sorumlu teknik servise olası yeni arızalar için en uygun yedek parçayı proaktif olarak önermeyi amaçlar. Sonuç olarak içinden geçtiğimiz COVID19 döneminde satış sonrası müşteri hizmetleri alanında hem ürün inovasyonu hem de süreç iyileştirmesi amaçlanmıştır.

Sınıflandırma algoritmaları perakende, finans, bankacılık, güvenlik, astronomi ve davranışsal ekoloji alanlarında yaygın olarak kullanılmasına (Kantardzic, 2011) ve dengeli sınıflı veri kümeleri için nispeten iyi performans göstermesine rağmen, karma veri tipi özelliğine sahip dengesiz sınıflı veri kümelerinin sınıflandırılması görece karmaşıktır (Liu vd., 2014). Bu açıdan, öngörülen yöntem hibrit bir sınıflandırma algoritmasına dayanmaktadır: Apriori algoritması, veri hazırlama aşamasında gözlemlenen veri anomalileri ve tutarsızlıklarını ele almak için uyarlanırken, C4.5 algoritmasıyla elde edilen öznitelik ağırlıkları, arıza kayıtları arasındaki benzerliğin hesaplanmasında ve kategorik özellikler üzerinde normleştirilmesinde etkin olarak kullanılır. Modelleme son adımı olarak, ağırlıklandırılmış k-En Yakın Komşuluk (kNN) algoritmasının iki adaptasyonu geliştirilmiştir. Sayımlı örnek tabanlı 1kNN (1kNNwC-instance-based kNN with count), özellik uzayının komşuluk bölgesinde baskın bir sınıfı temsil etme olasılığı daha yüksek olan büyük örneklem gruplarına daha fazla önem verirken, ortalama benzerlik değerine sahip eşgörünüm tabanlı 1kNN (1kNNwAS-instance-based kNN with average score), çekinik (veya aykırı) gözlemlerin ayırt edici gücünü dengelemeyi amaçlar. Önerilen yaklaşım, televizyon ürün grubuna CRISP-DM (Cross Industry Standard Procedure for Data Mining) referans modeli çerçevesinde uygulanmıştır.

Makale şu şekilde yapılandırılmıştır: Bölüm 2'de kNN uyarlamaları ile ilgili literatür gözden geçirmektedir. Bölüm 3, önerilen gerçek zamanlı yedek parça tahminleme çözümünü CRISP-DM referans modelinde geçen problem anlamlandırma, veri anlamlandırma, veri hazırlama ve modelleme aşamalarını göre tanıtmaktadır. Bölüm 4'de değerlendirme ve canlı geçiş aşamalarında gözlemlenen performansa göre deneysel sonuçlar tartışılmaktadır. Sonuç ve gelecek çalışmalar Bölüm 5'te özetlenmiştir.

2. LİTERATÜR TARAMASI

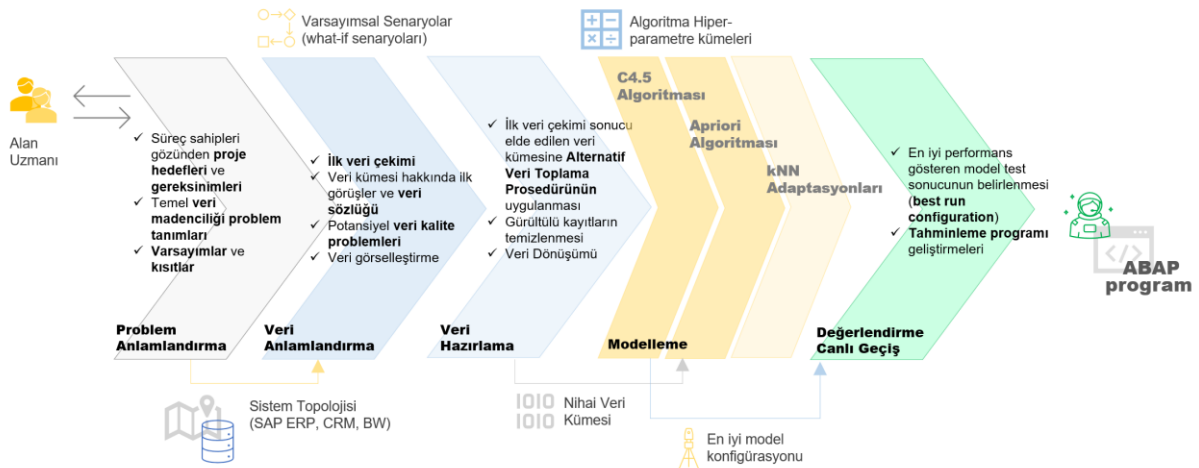
kNN algoritması, benzetim yoluyla sınıflandırmanın önemli bir temsilcisi olarak kabul edilir (Domingos, 2015). Bu çerçevede, en yakın komşu sayısını gösteren optimum k değerini bulmak, literatürde geçen çalışmaların çözmeye çalıştığı sorulardan biri olmuştur (Zhang vd., 2017; Zhu vd., 2016). Bu tür bir sınıflandırmada k değerini bulmanın yanı sıra, uzaklık hesabı da bir diğer meseledir. Ağırlıklı şema kullanımı ilk olarak (Dudani, 1976) tarafından tanıtılmıştır. Bu kNN varyantı, mesafe ağırlıklı k-en yakın komşu (DWkNN-Distance-Weighted k-Nearest Neighbor) olarak adlandırılır. (Tan, 2015), sınıfların dağılımına dayalı bir ağırlıklandırma stratejisi uygulayan komşu ağırlıklı k-en yakın komşu (NWkNN-Neighbor-Weighted k-Nearest Neighbor) algoritmasını önermektedir. (Mateos-Garcia vd., 2016)'de yapay sinir ağlarındaki yapıya benzer bir şekilde komşuluğun önemini belirterek ağırlıkları optimize eden bir teknik geliştirilirken, (Parvinnia vd., 2014) her eğitim gözlemi için eşleştirme stratejisine dayalı bir ağırlık hesaplaması önermiştir. Bu çalışmaya benzer olarak, (Aguilera vd., 2019) Newton'un yerçekimi kuvvetine dayanarak her gözleme ikili ağırlık (veya ilişki düzeyi) atanmasını önermiştir: kendi sınıfı (CC) ve farklı sınıf (CD) kavramlarına göre komşuluk sınırlarının belirlenmesi.

Standart kNN algoritmasının dengesiz sınıf dağılımının gözlemlendiği veri kümelerine uygulanması görece karmaşıktır. Bu nedenle, kENN (Yuxuan ve Zhang, 2011) ve CCW-kNN (Liu ve Chawla, 2011) uyarlamaları dengesiz veri sınıflandırmasında performansın iyileştirilmesi için önerilmiştir. kENN, pozitif eğitim örneklerinin tanımlandığı ve Gaussian toplarına genelleştirildiği bir eğitim aşaması önerirken, CCW-kNN, kNN algoritmasındaki prototipleri ağırlıklandırmak için sınıf etiketleri verilen özellik değerlerinin olasılığını kullanmaktadır. (Song vd., 2007), sorgu tabanlı bir mesafe hesaplamasına dayanan bir kNN sınıflayıcısı önermektedir. Bu sınıflayıcı, benzerlik ölçümünü yerel (LI-kNN) ve global (GI-kNN) olmak üzere iki açıdan ele alınmaktadır. Alternatif olarak, (Wang vd., 2011) kategorik özelliklerin hem dahili hem de öznitelikler arasındaki ilişkileri incelemek için birleştirilmiş bir nominal benzerlik sunmuştur. İlgili yaklaşımlar temel olarak dengeli sınıflı veri kümelerinde kümeleme esaslı uyarlamalara odaklanmıştır.

3. YÖNTEM

Yedek parça tahminleme senaryosu için ana uygulama yol haritası olarak CRISP-DM referans modeli izlenmiştir (Chapman vd., 1999). CRISP-DM referans modeli altı aşamadan oluşmaktadır: problem anlamlandırma, veri anlamlandırma, veri hazırlama, modelleme, değerlendirme ve canlı geçiş. Bu bölümde değerlendirme ve canlı geçiş hariç diğer aşamalar sırasıyla ele alınmıştır. Şekil 1'de öngörülen yaklaşım detaylıca gösterilmiştir.

Şekil 1. Öngörülen Yaklaşım.



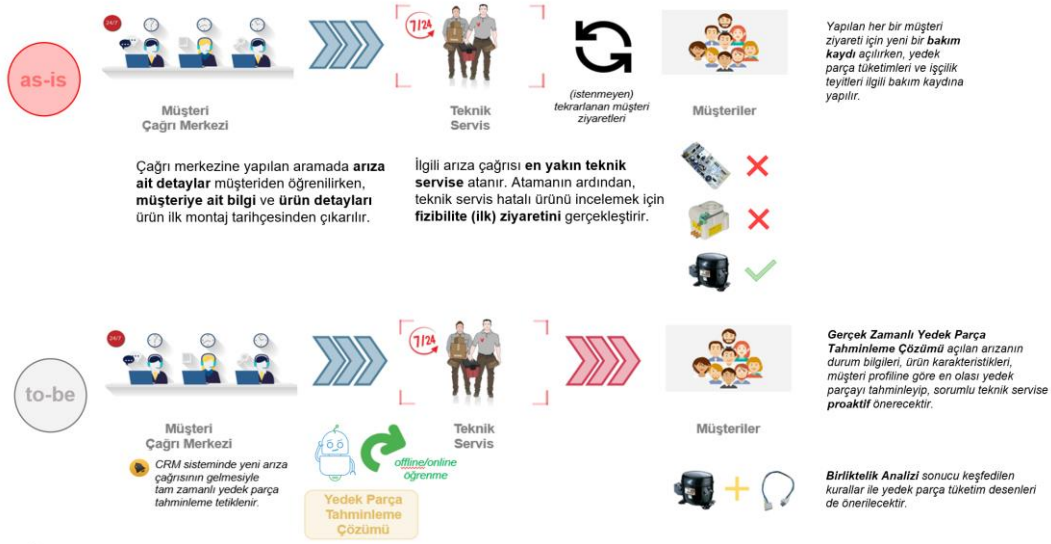
3.1. Problem Anlamlandırma

Mevcut durumda (as-is) müşteri çağrı merkezine yapılan her müşteri çağrısı, SAP CRM sisteminde yeni bir arıza kaydını tetikler. Bu görüşme sırasında müşteriden arızaya ait detaylar (örn. ürün grubu, en genelden en özele şikayet) öğrenilir. Görüşme sonlandıktan sonra, müşteri profili (örn. müşteri profili ve konumu) zenginleştirilirken, ürün bilgileri (örn. ürün kodu (SKU-stock keeping unit, malzeme tipi, malzeme grubu ve ürün hiyerarşisi) SAP CRM sisteminde yönetilen ürün montaj geçmişinden okunur. İzleyen adım olarak, ilgili arıza çağrısı müşterinin bulunduğu lokasyona göre en yakın teknik servise atanır. Son olarak teknik servis, arıza nedenini ve arızalı bileşeni kontrol etmek için bir fizibilite ziyareti yapar. İlgili arıza için gerçekleşen her bir müşteri (saha) ziyareti tekil bir bakım onarım kalemiyle takip edilip bu müşteri ziyaretlerindeki yedek parça tüketimleri veya bakım aktiviteleri ilgili bakım onarım kalemine teyit edilir.

Öngörülen çözüm (to-be) olarak, ilgili arıza çağrısı için en olası yedek parçayı tahminleyen ve bu tahmini gerçek zamanlı olarak teknik servise ileten bir yedek parça tahminleme çözümünün mevcut sürecin içine konumlandırılması amaçlanmaktadır. Böylece teknik servis, günlük araç içi yedek parça stoğunu ve günlük müşteri rotalarını proaktif olarak yeniden düzenleyebilir. Ayrıca, arıza çağrısı başına ortalama müşteri ziyareti sayısının ciddi oranda azaltılması hedeflenmektedir. Bu açıdan, operasyonel satış maliyetleri ve yedek parça tüketim giderleri müşteri ziyaretiyle güçlü bir şekilde ilişkilidir. İlgili

anahtar performans göstergesindeki olası düşüşler, gelir tablosundaki ilgili gider kalemlerini de azaltacaktır (örn. taşıma maliyetleri, teknik servis araçlarının bakım ve amortisman maliyetleri vb.). Satış sonrası hizmet kalite seviyesinin yükseltilmesi ve sıkça tekrarlanan çözümsüz müşteri ziyaretlerinden zarar gören kurumsal marka değerinin uzun vadede iyileştirilmesi de dolaylı kazançlar olarak ele alınabilir. Mevcut ve öngörülen durumlar Şekil 2'de gösterilmektedir.

Şekil 2. Mevcut ve Önerilen Durumlar.



3.2. Veri Anlamlandırma

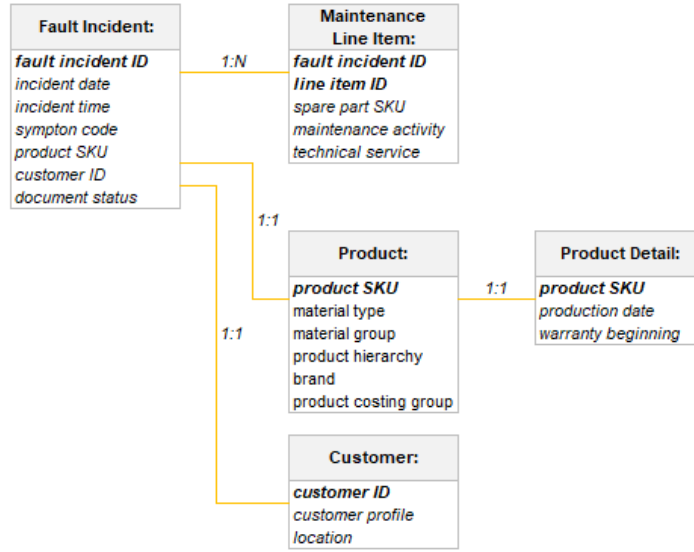
Veri anlamlandırma aşaması, başlıca veri kaynaklarının belirlenmesi, bu veri kaynakları arasındaki entegrasyon ve ilk ham veri kümesini oluşturan başlıca özelliklerin tanımlanması gibi operasyonları kapsar. İlk olarak, ilk ham veri kümesinin temel özelliklerini listeleyen veri sözlüğü tanımlanır. Sonrasında, öznitelikler ve ilk ham veri kümesindeki olası kalite problemlerinin iyileştirilmesi için veri keşfi gerçekleştirilir.

3.2.1. Veri Tanımlama

İlgili iş senaryosu beş veri kaynağından oluşur. Şekil 3'te verilen bağlam diyagramında veri kaynakları arasındaki ilişkiler gösterilmektedir.

- Arıza çağrısı (fault incident). Arıza çağrısı, ilgili arıza kaydının başlık bilgilerini (örn. arıza tanıtıcısı, olay tarihi ve saati, arıza semptom kodları, belge durumu, ilgili müşteri tanımlayıcısı ve ürün kodu) tutar.
- Bakım onarım kalemi (maintenance line item). Bakım onarım kalemi, ilgili müşteri ziyaretindeki yedek parça tüketimini ve bakım onarım aktivitesini tutar. Arıza çağrısı ile bakım onarım kalemi arasında bire çok (1:N) ilişki vardır.
- Ürün (product). Ürün, ürün kodu, malzeme türü, malzeme grubu, ürün hiyerarşisi, marka ve ürün maliyet grubu gibi arızalı ürünle ilgili temel özellikleri içerir.
- Ürün detayları (product detail). Ürün detayları, üretim tarihi ve garanti başlangıç tarihi gibi önemli üretim ayrıntılarını içerir. Ürün ve ürün detay veri kaynakları arasında bire bir (1:1) ilişki vardır.
- Müşteri (customer). Müşteri, müşteri profilini ve konumunu şehir ve bölge cinsinden tutar.

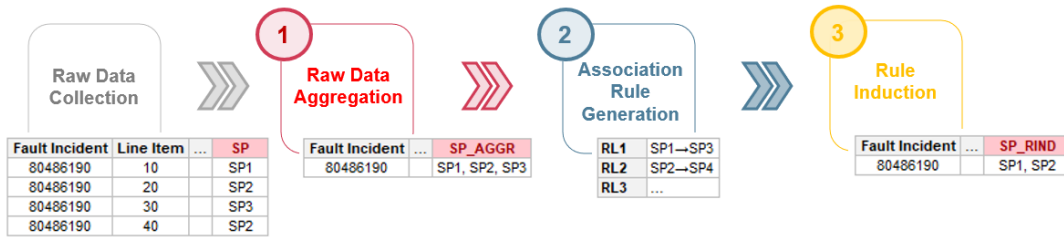
Şekil 3. Veri Sözlüğü Bağlam Diyagramı.



3.2.2. Veri Keşfi

Veri keşfi adımı, nitelikler arasındaki korelasyonu değerlendirirken, olası veri anomalileri veya tekrarının (redundancy) oluşup oluşmadığını kontrol eder. Bu açıdan, ilk ham veri kümesindeki her bir gözlem benzersiz bir bakım onarım kalemiyle temsil edilirken, yedek parça hedef sınıfı (target class) hariç, tüm geçerli özniteliklerin arıza çağrısı, ürün, ürün detayı ve müşteri gibi aynı veri kaynaklarından elde edilmesi sonucu önemli bir veri tekrarı sorunu ortaya çıkmıştır. Sonuç olarak, ilk ham veri toplamada tekrarlayan öznitelik değerleri ve farklı yedek parça hedef değeri birleşimi sonucu kirli arıza kayıtları ortaya çıkmaktadır. Çözüm olarak, her bir gözlem (instance) benzersiz bir arıza kaydıyla ilişkilendirilerek daha yüksek bir soyutlama düzeyinde (abstraction level) veri kümesinin yönetilmesine karar verilmiştir. Bu nedenle, Şekil 4'te gösterildiği gibi alternatif veri toplama prosedürü geliştirilmiştir.

Şekil 4. Alternatif Veri Toplama Prosedürü. Dört bakım onarım kalemine sahip arıza kaydı prosedür sonucunda iki yedek parça koduna sahip tek bir arıza kaydına dönüştürülmüştür.



Alternatif veri toplama prosedürü aşağıdaki üç temel adımdan oluşmaktadır:

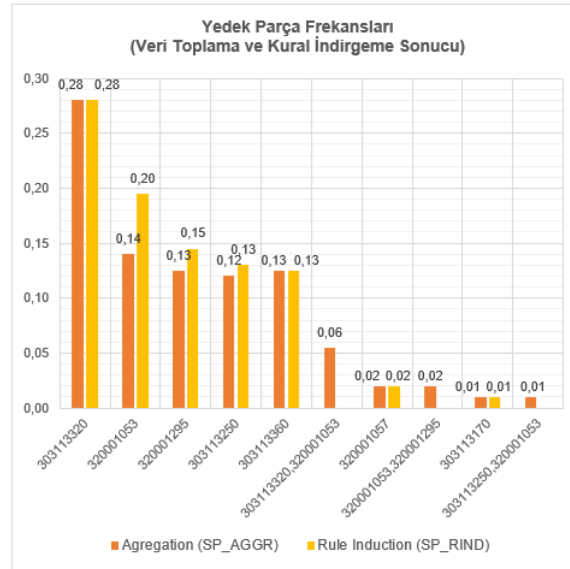
- Ham veri toplama (raw data aggregation). Bu ilk adım, aynı arıza olayıyla ilgili her bir bakım onarım kaleminde geçen yedek parça tüketim kodlarını (SP) toplar. Bu açıdan, ilgili bakım onarım kalemlerindeki yedek parça değerleri aktarılıp, değerler türetilmiş tek bir hedef sınıf alanında (SP_AGGR) birleştirilir. Yedek parça değerleri birleştirilirken, yinelenen değerler tekilleştirilip benzersiz değerler artan düzende sıralanır, örn. <SP1, SP1, SP3, SP2> yedek parça torba değerleri (bag) ham veri toplama adımı sonucu <SP1, SP2, SP3> sırasına (sequence) dönüşür.
- Birliktelik kuralı oluşturma (association rule generation). Elde edilen toplu ham veri kümesi sık görülen öge kümelerine benzemekte olup, bu öge kümeleri temel değişkenlere ilişkin bir Boolean yedek parça değeri vektörü ile temsil edilebilir. Oluşan vektörler, sıklıkla tekrarlayan yedek parça birlikte tüketim örüntülerinin analizi için kullanılabilir. Bu modeller birliktelik

kuralları (association rules) şeklinde temsil edilebilir. Apriori, boolean birliktelik kuralları cinsinden sık kullanılan öğe kümeleri için önerilen bir algoritmadır. İlgili algoritma sık kullanılan öğe kümesi özelliklerine ilişkin ön bilgileri kullanmasına dayanmaktadır (Kantardzic, 2011). Bu bağlamda gerçek zamanlı yedek parça tahminleme çözümünde *arules* ve *arulesViz* R paketleri kullanılmıştır. Apriori algoritmasıyla elde edilen ilişkilendirme kuralları *min_lift* ölçütüne göre filtrelendir; *min_lift* > 1.0 eşiği yalnızca pozitif ilişkili yedek parça kombinasyonlarının çıkarılması için kullanılır.

- Kural indirgeme (rule induction). Kural indirgeme adımı, toplu yedek parça hedef sınıf değerlerini (SP_AGGR) birliktelik kurallarına göre rafine edilmiş formlara dönüştürür. Bu açıdan anlamlı ve görece güçlü birliktelik kuralları *min_confidence* ölçütü ile belirlenir. Filtrelenen birliktelik kuralları, güven (confidence) değerlerine göre sıralanır. Daha sonra, toplu ham veri kümesindeki her bir yedek parça hedef sınıf değerinde (SP_AGGR), birliktelik kuralına ait ön koşul (antecedent) ve sonuç (consequent) değerlerinin mevcut olup olmadığı kontrol edilir. Mevcut olması durumunda, sonuç değeri toplu yedek parça hedef sınıf değerinden çıkarılıp elde edilen yeni değer kural indirgenmiş hedef sınıf özniteliğine (SP_RIND) işlenir. Aksi durumda, toplu yedek parçanın orijinal değeri, kurala indirgenmiş yedek parça hedef sınıfına kopyalanır.

Şekil 5'te alternatif veri toplama prosedürü sonucu toplu ve kural indirgenmiş hedef sınıf frekanslarındaki geçişmeler görselleştirilmiştir. Örneğin, <320001053>→<303113320> birliktelik kuralı nedeniyle 320001053 frekansında önemli bir artış meydana gelmektedir.

Şekil 5. Alternatif Veri Toplama Prosedürü Sonuçları.

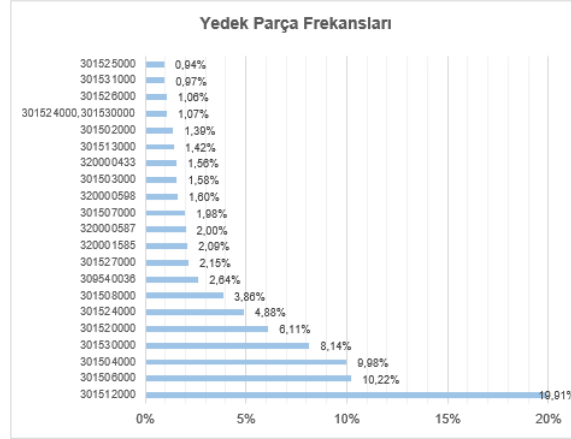


3.3. Veri Hazırlama

Veri hazırlama, ilk ham veri kümesinin nihai haline dönüştürülmesi için gerekli olan veri entegrasyonu, dönüştürme ve temizleme operasyonlarını kapsar. Veri entegrasyonu adımı, televizyon ürün grubuna ait 2014-2018 yılları arasında yaratılan arıza kayıtlarının CRM ve SAP ERP kaynak bilgi sistemlerinden çıkarılması için Şekil 3'te verilen bağlam şemasına göre tasarlanan SQL scripti yürütülür. Ardından, bakım onarım kalemi düzeyinde gözlemlenen anormallikler ve veri tekrarlarını önlemek için Bölüm 3.2.2'de bahsedilen alternatif veri araştırma prosedürü uygulanır. Ham veri toplama adımı sonucu dikey bir veri indirgemesi meydana gelir. Örneğin, 750K bakım onarım kalemi veri hacmi alternatif veri toplama prosedürü sonucu 350K seviyelerine indirgenmiştir. Buna ek olarak, hiyerarşik özniteliklere (örn. müşteri konumu, ürün hiyerarşisi ve şikayet kodları) bağlı öznitelik tekrarından kaçınmak için nispeten daha yüksek ayrıntı düzeyine ve daha geniş değer aralığına sahip öznitelikler seçilir. Veri dönüştürme adımı ise, ay ve yıl öznitelikleri, temel alınan tarih tipi özniteliklerden (örn. arıza tarihi ve üretim tarihi vb.) ayrıştırılır. Ayrıca ürün yaşı, ürün stok yaşı ve garanti durumu gibi yeni özellikler de türetilmektedir.

Kural indirgeme işlemi sonucu, geçerli hedef sınıf (SP_RIND) değer aralığı 1267 farklı üyeden oluşmaktaydı. Bu nedenle, frekansı görece düşük yedek parça kodları veri temizleme adımında ihmal edilip bu işlem sonucu veri kümesinde toplam %95.27'lik bir kapsama elde edilmiştir. Ayrıca, ALL, TOPN (örn. frekans değeri en yüksek N yedek parça kodu) gibi frekans sırasına göre çeşitli yedek parça grupları tanımlanmıştır. Şekil 6'da verilen frekans histogramına göre sınıf dengesiz veri kümesi durumu vurgulanabilir.

Şekil 6. Veri Hazırlama Sonuçları.



3.4. Modelleme

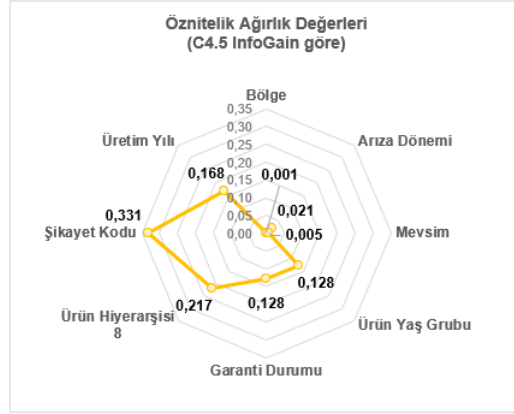
Yedek parça tahminleme iş senaryosu, kural indirgenmiş hedef sınıf özniteliğinin (SP_RIND) varlığından dolayı bir tür denetimli öğrenmedir (supervised learning) ve bu senaryonun ana amacı, yedek parça tüketim fenomeninin altında yatan en güçlü iç görü ve kuralları keşfetmektir. Veri sözlüğüne göre, nihai veri kümesindeki hemen hemen tüm öznitelikler, geniş bir değer aralığına kategorik tipli veri yapısına sahiptir. Bu nedenle, hedef sınıf dengesiz karma veri tiplerine sahip nihai veri kümesi için C4.5 ve Apriori algoritmalarını ağırlıklı kNN varyasyonlarıyla birleştiren hibrit bir model tasarlanmıştır.

3.4.1. C4.5

C4.5, karar ağaçlarının yukarıdan aşağıya özyinelemeli böl-ve-yönet (command-and-conquer) tarzında sınıflandırıcı olarak oluşturulduğu açgözlü ve geriye dönük olmayan bir yaklaşımı benimser (Kantardzic, 2011). Geçerli öznitelik seçim yöntemi olarak, ilgili veri kümesini hedef sınıfa göre en iyi ayırt eden özniteliği seçmek için buluşsal (heuristic) prosedürler uygulanır.

Yedek parça tahminleme senaryosu kapsamında, bilgi kazancı (infogain) öznitelik seçimi ve *min_split* hiper-parametre değeri 50 olarak belirlenip *rpart* R paketi uygulanmıştır. Sonuç olarak %42.5'lik doğruluk (accuracy) performansı aday modeller için kesin referans (ground truth) teşkil ederken, Şekil 7'de gösterildiği gibi yedek parçaların tahminlemede öznitelik önem ağırlıkları da elde edilmiştir. Her ne kadar Jaccard (Pang-Ning et al., 2006), kosinüs benzerlik (Liu et al., 2014) ve Goodall benzerlik (Boriah, 2008) gibi ölçütler kategorik özniteliklerle kullanılsa da, bu ölçütler hedef sınıf değerleriyle birlikte ortaya çıkan bilgi katma değerini ihmal edilir. Bu nedenle, C4.5 algoritmasıyla elde edilen öznitelik ağırlıkları, Bölüm 3.4.3'te bahsedilecek kNN uyarlamalarında boyutlar (yada öznitelikler) arası ağırlıklandırma (*interDim_weight*) için esas teşkil eder.

Şekil 7. C4.5 Algoritmasına göre Özniteliklerin Önem Ağırlığı. Görece, ürün hiyerarşisi8 ve şikayet kodu, yedek parça tahminleme senaryosunda ana etmenlerdir.



3.4.2. Apriori

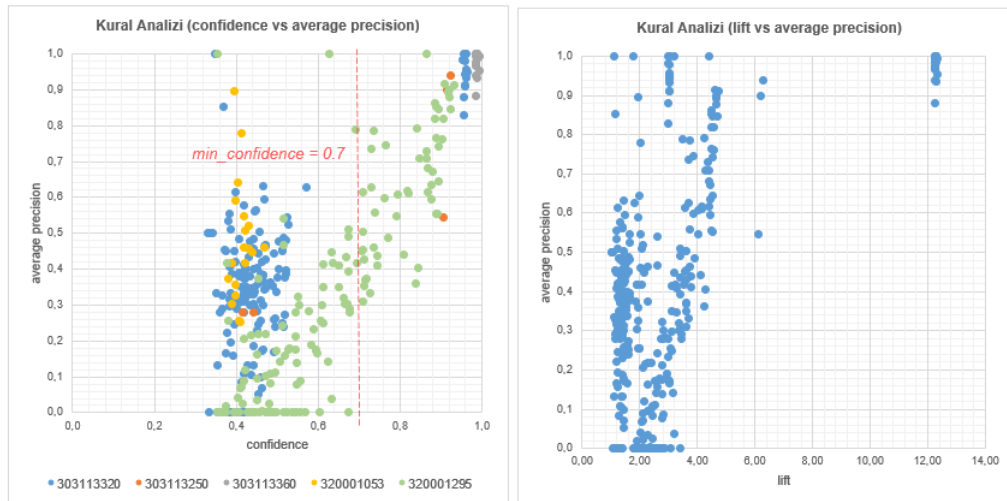
Bölüm 3.2'de vurgulandığı gibi, Apriori algoritması temel olarak yedek parça kodları arasındaki önemli birlikte tüketim kurallarını araştırır. Bunun yanında, somut bir sınıflandırıcı olmaksızın hedef sınıfı tahmin etmek için IF/THEN formatında farklı uzunluklara sahip kurallar da türetebilir. Bu açıdan, son veri kümesi için $min_support > 0.01$ koşuluyla Apriori algoritması uygulanıp yaklaşık 2710 kuralı üretilmiştir. Tablo 1'de üretilen bazı Apriori nedensellik kuralları verilmiştir.

Tablo 1. Apriori ile oluşturulan IF/THEN Tipi Kural Örnekleri.

Ön Koşul (IF)	Sonuç	Support	Confidence	Lift
ZPRDAGE_GR=G1,ZSIKAYET_T=S949285	303113250	0,0100	0,9248	6,2660
ZGRNTDHL=Y,ZSIKAYET_T=S949285	303113250	0,0225	0,9147	6,1976
MEVSIM=SONBAHAR,ZPRDAGE_GR=G0	303113320	0,0326	0,4034	1,2671
ZPRDAGE_GR=G0,ZURTMONTH_YR_T=Y2015	303113320	0,0284	0,4155	1,3053
BOLGE=MARMARA,ZPRDAGE_GR=G0	303113320	0,0306	0,4508	1,4161

$min_lift > 1.0$ özelliği olan kurallar, farklı arıza yılı, yedek parça grupları ve doğrulama yöntemleri ile tasarlanmış 66 test senaryosuna göre doğrulanmıştır. ALL yıl grubunun ortalama doğruluğu (average accuracy) %52 iken, 2017 yılı için doğruluk (average accuracy) %55.7 ile zirve yapmaktadır. En düşük ortalama doğruluk %45.3 ile 2015 yılında elde edilmiştir. Ayrıca, Şekil 8'de gösterildiği gibi kural kesinliği (precision) ile güven (confidence) arasında anlamlı bir ilişki gözlemlenmektedir. Ancak, güven değeri [0.4, 0.6] aralığında olan kurallar ortalama %25.27 kesinlik ile ciddi oranda yanlış tahminlere (FP-false positive) yol açmıştır.

Şekil 8. Apriori Kural Analizi (güven vs. kural kesinliği).

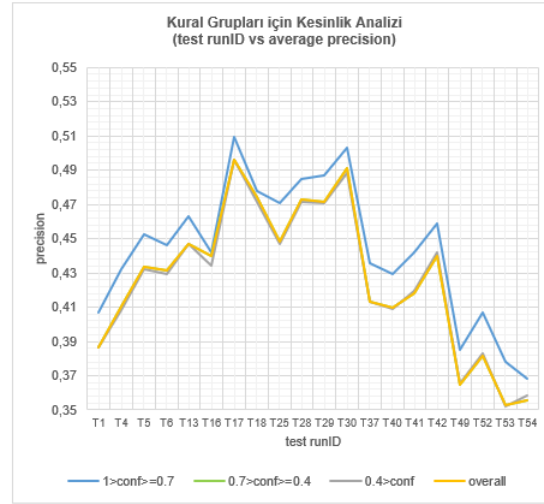


Tablo 2'ye göre $min_confidence = 0.7$ dirsek noktası görece zayıf kuralların filtrelenmesi için iyi bir başlangıçtır. Şekil 9'da kural grupları arasındaki ortalama kesinlik farkları vurgulanmaktadır.

Tablo 2. Güven Dirsek Noktasının Belirlenmesi.

	confidence dirsek noktası				
	0,60	0,70	0,80	0,90	1,00
avg precision	67,26%	76,05%	84,20%	93,27%	0,00%
kural sayısı	115	90	68	44	0

Şekil 9. Kural Grupları için Ortalama Kesinlik Değerleri.



3.4.3. kNN Uyarlamaları

kNN sınıflandırması, belirli bir gözlem kümesini, ona benzer veri havuzunda eğitim için ayrıştırılan kayıtlarla karşılaştırarak benzetme yoluyla çevrimiçi öğrenme (online learning) şemasına dayanmaktadır (Kantardzic, 2011). Eğitim kayıtları, n -boyutlu bir uzayda temsil edilirken, yeni bir gözlem geldiğinde, kNN sınıflandırıcısı ilgili gözleme en yakın olan k kaydı araştırır. Tipik olarak, her bir özelliğin değerleri, mesafe hesaplamasından önce normalleştirilmelidir. Ancak, kategorik öznitelikler için mesafe hesaplaması ve bu özniteliklerin kendi değer aralığında yer alan münferit üyeler arasındaki görece benzerliğinin ölçülmesi yedek parça tahmin senaryosunda ortaya çıkan iki ana meseledir. kNN algoritmasında ilk uyarılma olarak, temeldeki yakınlık ölçümü, Denklem 1'de gösterildiği gibi bir benzerlik ölçümüne, $simScr(insi, insj)$, dönüştürülmüştür.

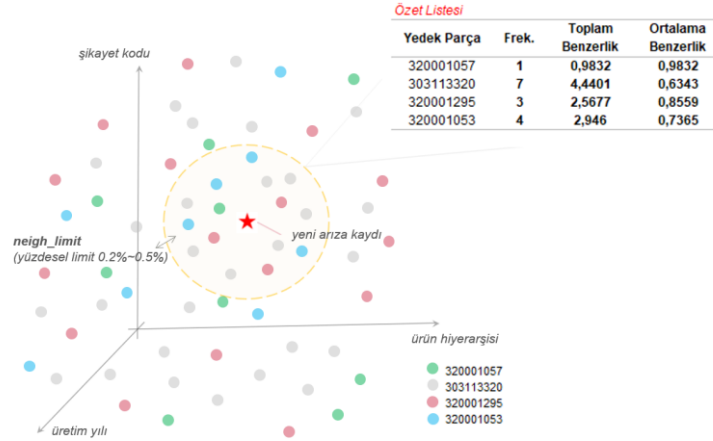
$$simScr(insi, insj) = \sum_{dim=1}^n interDim_weight_{dim} \times intraDim_weight(insi, dim, insj, dim) \quad (1)$$

Denklem 1'de, $interDim_weight$, her bir anlamlı özniteliğe (yada boyuta) tayin edilen normalleştirme ağırlığını temsil eder ve C4.5 algoritmasıyla elde edilen öznitelik ağırlığı kategorik özniteliklerin normalleştirilmesine uyarlanmıştır. $intraDim_weight$ faktörü ise, hiyerarşik (sıralı) özniteliklerin farklı seviyeleri arasındaki benzerlik derecesine karşılık gelir. Benzerlik ölçümü bağlamında iki farklı kNN adaptasyonu geliştirilmiştir: örnek tabanlı kNN (IkNN-instance based kNN) ve ortalama kNN (AkNN-average kNN).

IkNN adaptasyonunda, nihai veri kümesindeki arıza kayıtları ile yeni arıza gözlemi arasında benzerlik Denklem 1'e göre hesaplanır. Hesaplanan değerlere göre en yakın komşu arıza kayıtları komşuluk limiti ($neigh_limit$) hiper-parametresine göre belirlenir. Bu parametre, küre benzeri bir bölgenin içinde yer alan en yakın arıza kayıtlarını belirleyen yüzdelik bir sınırdır. Daha sonra Şekil 10'da gösterildiği gibi, yedek parça hedef sınıf değerlerine göre frekans, toplam ve ortalama benzerlik değerleri tahmin sonuç listesinde özetlenir. Bu noktada IkNN'nin iki varyasyonu uygulanmaktadır:

- Sayımlı örnek tabanlı kNN (IkNNwC-instance based kNN with count). Yedek parça tahmini sonuç listesindeki yedek parçaların *frekans* değerine göre belirlenir. Bu sayede ilgili komşuluk bölgesinde yer alan farklı bir sınıfı temsil etme olasılığı daha yüksek olan nesnelere daha az önem verilmesi amaçlanmaktadır. Başka bir deyişle, aykırı (yada harici) sınıf değerlerine karşı daha sağlam bir tahminleme modeli ortaya koymaktır.
- Ortalama benzerlik tabanlı kNN (IkNNwAS-instance based kNN with average score). Yedek parça tahmini sonuç listesindeki yedek parçaların *ortalama benzerlik* değerine göre belirlenir. Bu varyasyon, daha az temsil gücü olan minör sınıflara daha fazla temsil şansı vererek sınıf dengesiz veri kümelerinde daha iyi tahminleme performansına imkan sağlar.

Şekil 10. Yeni Bir Arıza için IkNN Algoritmasına göre Tahminleme.



Son olarak *k-limit* hiper-parametresine göre yedek parça tahmin sonuç listesi aşağıdaki prosedüre göre sonuçlandırılır. Tablo 3'te IkNN adaptasyonlarında *k-limit* uygulaması örneklenmiştir.

- *k-limit* = n (örn. n = 1, 2...) durumunda, frekans yada ortalama benzerlik sıralamasına göre en üstteki n yedek parça değeri tahmin olarak süreç sahibine döndürülür. İlgili iş senaryosunda n için azami değer 2 ile sınırlanmıştır.
- *Dinamik k* (örn. DK:[0, 1]) durumunda, frekans yada ortalama benzerlik değerleri için standart sapma hesaplanıp, ardışık iki yedek parça değeri arasındaki fark $DK \times \text{stdDev}$ değerinden büyükse, kontrol edilen tüm yedek parça değerlerinin birleşimi süreç sahibine tahmin olarak döndürülür. Aksi durumda tahmin sonuç listesinde kalan satırlar kontrol edilmeye devam edilir. Özellikle düşük standart sapma ve büyük DK değerleri ilgili modelin seçici tahminleme yeteneğini köreltir. Sonuç olarak, daha uzun tahmin listesi tahminleme performansının düşmesine neden olur.

Tablo 3. IkNN Adaptasyonlarında k-limit Uygulama Örnekleri.

Instance based KNN with count			dynamic k					
Yedek Parça	Frekans	ardışık fark	k=1	k=2	DK=1.0	DK=0.75	DK=0.5	DK=0.25
303113320	7		√	√	√	√	√	√
320001053	4	3		√				
320001295	3	1						
320001057	1	2						
stdDev:2.165								
Instance based KNN with average similarity			dynamic k					
Yedek Parça	Ortalama Benzerlik	ardışık fark	k=1	k=2	DK=1.0	DK=0.75	DK=0.5	DK=0.25
303113320	0,9832		√	√	√	√	√	√
320001053	0,8559	0,1273		√	√	√	√	√
320001295	0,7365	0,1194			√	√	√	√
320001057	0,6343	0,1022			√	√	√	√
stdDev:0.1305								

note: √ işareti ilgili yedek parça kodunun tahmin listesine dahil edilmesi anlamına gelir.

4. DENEY SONUÇLARI VE BULGULAR

4.1. Değerlendirme

Bölüm 3.4.2'de vurgulandığı gibi, [0.4, 0.6] aralığında güven değerine sahip Apriori kuralları, ortalama %25.27 kesinliğe sahip olarak hatalı tahminler yapmaya eğilimindedir. Bu sebeple, $min_confidence = 0.7$ dirsek değeri belirlenerek aşağıdaki hibrit (ensemble) sınıflandırma algoritması uygulanmıştır:

- İlk olarak, nispeten güçlü Apriori kuralları ile yeni arıza kaydı tahmin edilmeye çalışılır.
- Apriori kuralları ile herhangi bir tahminleme yapılamazsa, çevrimiçi öğrenme mantığına göre kNN adaptasyonları uygulanır.

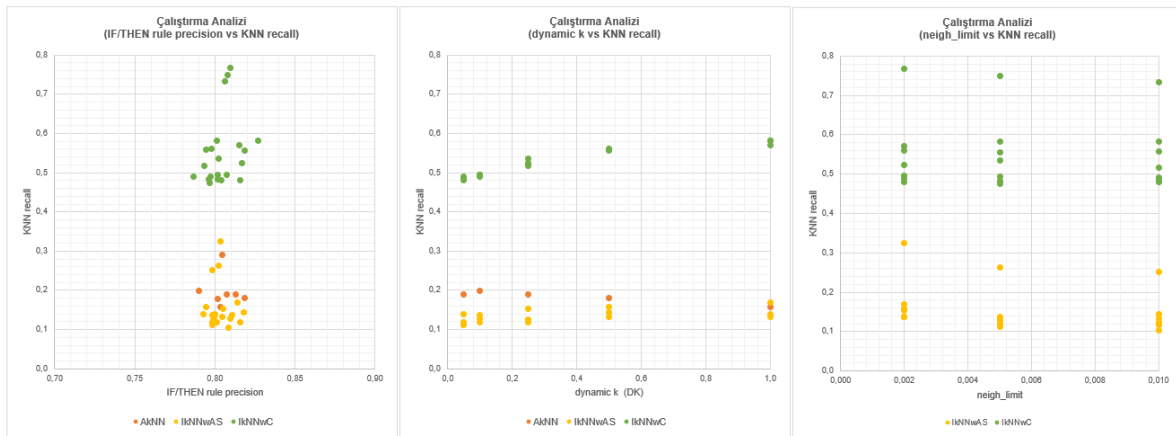
Bu nedenle, farklı arıza yılları, yedek parça grupları, kNN adaptasyonları, hiper-parametre (örn. $neigh_limit$ ve k_limit) konfigürasyonu ve doğrulama yöntemlerine göre 56 test senaryosu tasarlanmıştır. Tablo 4'te gösterildiği gibi, saf Apriori kurallarının ortalama doğruluğu (average accuracy) %54.02 iken, Apriori ve 1kNN adaptasyonu kombinasyonu ortalama doğruluğu (average accuracy) %77.95 seviyesine yükselmiştir. TOP3 yedek parça grubu için performans yaklaşık %85.2 iken, TOP6 grubu için %79,24 seviyesindedir.

Tablo 4. Farklı Algoritma Kombinasyonları ve Yedek Parça Grupları için Ortalama Doğruluk Değerleri.

Apriori				Apriori (min_confidence=0.7) ve AkNN Kombinasyonu				Apriori (min_confidence=0.7) ve 1kNN			
test sayısı	TOP3	TOP6	ALL	test sayısı	TOP3	TOP6	ALL	test sayısı	TOP3	TOP6	ALL
60	59,86%	55,68%	54,02%	7	38,06%	43,01%	43,46%	49	85,20%	79,24%	77,95%

Şekil 11'de verilen çalıştırma analizine göre, [0.78, 0.82] kesinlik aralığındaki Apriori kuralları yoğun olarak tahminlemede kullanılırken, 1kNNwC varyasyonu görece daha iyi bir tahminleme performansına sahiptir. Dinamik-k (DK) hiper-parametresindeki artışlar, 1kNNwC varyasyonunun duyarlılık (recall) değerlerinde olumlu bir etkiye sahip olsa da, AkNN ve 1kNNwAS varyasyonları için ters etki geçerlidir. Bunun nedeni, 1kNNwC varyasyonu komşu bölge sınırlarında yapılan değişikliklere karşı daha sağlam olmasına rağmen, 1kNNwAS varyasyonunun seçici tahminleme yeteneklerinin bu değişikliklere karşı daha kırılgandır. Bu nedenle, minör (yada frekansı görece düşük) sınıfların ayırt ediciliği kaybolur.

Şekil 11. kNN Adaptasyonları için Çalıştırma Analizi.

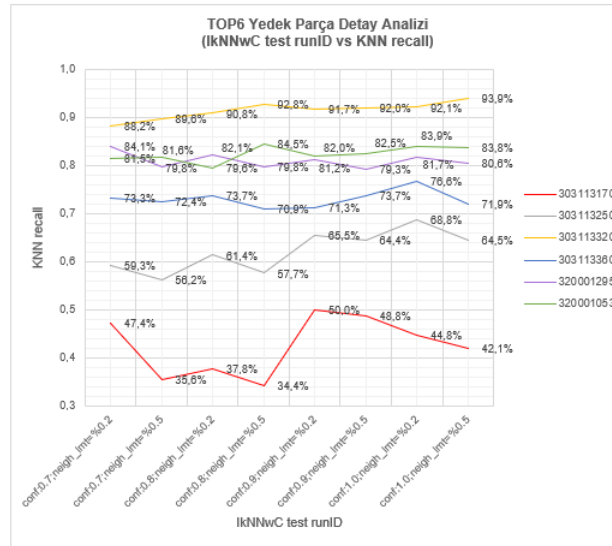


Bir sonraki adım olarak, Bölüm 3.4.2'de belirtilen $min_confidence$ dirsek noktası [0.7, 1.0] aralığında doğrusal olarak artırılır. Sonuç olarak, $min_confidence = 1.0$ koşuluyla filtrelenen Apriori kuralları ve 1kNNwC varyasyonu ($k_limit = 2$ ve $neigh_limit = \%0.2$) kombinasyonu Tablo 5'te gösterildiği gibi %80.68'lik bir ortalama doğruluğa (average accuracy) ulaşır.

Tablo 5. Değişen Güven Değeri, Farklı Algoritma Kombinasyonları ve Yedek Parça Grupları için Ortalama Doğruluk Değerleri.

	yedek parça grupları		
	ALL	TOP3	TOP6
Apriori	54,02%	59,86%	55,68%
Apriori (min_conf:0.7) + 1kNNwC	77,93%	85,20%	79,24%
Apriori (min_conf:0.8) + 1kNNwC	77,99%	85,25%	79,88%
Apriori (min_conf:0.9) + 1kNNwC	78,98%	86,13%	80,47%
Apriori (min_conf:1.0) + 1kNNwC	80,68%	87,66%	82,24%

Şekil 12’de gösterildiği gibi, yedek parça detayında duyarlılık (recall) değerleri münferit incelendiğinde 303113250 (yaklaşık frekans %13.01) yedek parça kodu, duyarlılık (recall) değerinde yaklaşık %7.8 gibi önemli bir artışa sahiptir. Benzer eğilim, en yüksek frekans değerine sahip 303113320 yedek parça kodu için de geçerlidir. İlgili yedek parçanın duyarlılık (recall) değerlerindeki %2.8’lik bir artış, Tablo 5’te gösterildiği gibi doğruluk üzerinde önemli bir kaldıraç etkisine sahiptir.

Şekil 12. TOP6 Yedek Parça Grubu için Farklı Algoritma Kombinasyonlarına göre Duyarlılık Değerleri.

4.2. Canlı Geçiş

Değerlendirme aşamasında elde edilen performans sonuçlarına göre, en iyi modelleme konfigürasyonu 1kNNwC varyasyonu (k -limit = 2 ve $neigh_limit$ = %0.2) Apriori kuralları ($min_confidence$ = 1.0) kombinasyonu ile elde belirlenmiştir. İlgili hibrit sınıflandırma modeli, Şekil 13’te gösterildiği gibi SAP sisteminde müşteri tanımlı bir tahminleme işlemi olarak canlı kullanıma alınmıştır. Fonksiyonel olmayan performans ölçümüne göre, yeni bir arıza kaydının tahminleme işlem süresi yaklaşık 7.79 saniyedir. Ardışık iki arıza kaydı arasındaki ortalama süre yaklaşık 36.5 saniye olduğundan, gerçek zamanlı yedek parça tahminleme teknik olarak mümkündür.

Şekil 13. Yedek Parça Tahminleme İşlemi Örnek Sonuç.

Yedek Parça Tahminleme Test Aracı							
Model: MD001-Eylül Televizyon Demo Tarih: 20.10.2021 Kullanıcı: EESGIN Kayıt Sayısı: 3 Sonuç Sayısı: 7 İşlem Süresi: 1							
Arıza Çağrısı	YP MG	Uzun Metin	Değer	Bir Tük YP	Uzun Metin	Kural Değeri	Servis Yorum
8057434524	303113360	TV-UJ/K	0,9687500			0,0000000	YP ihtimal çok yüksek + yerinde çözüm çok kolay
	303113320	TV-Şase	0,0312500			0,0000000	YP ihtimal çok düşük + yerinde çözüm zor
8057434524			1,0000000			0,0000000	
8057892465	320001053	TV-Besleme Kartı	0,6363636			0,0000000	YP ihtimal yüksek + yerinde çözüm çok zor
	320001295	TV-Led Bar Grubu	0,3636364			0,0000000	YP ihtimal normal + yerinde çözüm normal
	320001053	TV-Besleme Kartı	0,0000000	303113320	TV-Şase	0,2432000	
8057892465			1,0000000			0,2432000	
8058466061	303113320	TV-Şase	0,8620690			0,0000000	YP ihtimal çok yüksek + yerinde çözüm zor
	303113170	TV-Hoparlör/Mikrofon	0,1379310			0,0000000	YP ihtimal düşük + yerinde çözüm çok kolay
8058466061			1,0000000			0,0000000	
			3,0000000			0,2432000	

5. TARTIŞMA VE SONUÇ

Bu makale, yedek parça tahminleme iş senaryosu için hibrit bir sınıflandırma algoritması önermektedir. Apriori algoritması veri keşfinde ortaya çıkan anomalileri ve veri tekrarlarını ortadan kaldırırken, C4.5 algoritmasıyla elde edilen öznitelik ağırlıkları arıza kayıtları arasındaki komşulukların hesaplanmasında kullanılmıştır. Son olarak, iki ağırlıklandırılmış kNN varyasyonu uygulanmıştır: İkNNwC, öznitelik uzayı komşuluk bölgesinde baskın sınıfı temsil etmesi muhtemel olan yüksek frekanslı gözlemlere önem verirken, İkNNwAS minör sınıfların ayırt edicilik gücünü dengelemeyi amaçlamıştır.

DeneySEL sonuçlara göre önerilen hibrit sınıflandırma algoritması %40 seviyelerinde olan yedek parça tahminleme referans performansını ikiye katlamaktadır. Bu performans, arıza başına ortalama müşteri ziyareti sayısında yarı yarıya düşmesi anlamına gelirken, özellikle satış operasyon giderlerinde önemli bir iyileşme öngörülmektedir. Gelecek çalışmalar kapsamında, ilgili yedek parça tahminleme iş senaryosunun diğer ürün gruplarına da uygulanması hedeflenmektedir.

TEŞEKKÜRLER

Gerçek Zamanlı Yedek Parça çözümü KOSGEB ArGe ve İnovasyon Destek Programı çerçevesinde 3N2L8 proje koduyla desteklenmektedir.

KAYNAKÇA

- Aguilera, J., Gonzalez, L. C., Montes-y-Gomez, M., ve Rosso, P. (2018). A new weighted k-nearest neighbor algorithm based on newton's gravitational force. In *Iberoamerican Congress on Pattern Recognition*, 305-313. doi:10.1007/978-3-030-13469-3_36.
- Boriah, S., Chandola, V., ve Kumar, V. (2008). Similarity measures for categorical data: A comparative evaluation, In *Proceedings of the 2008 SIAM international conference on data mining*, 243-254. doi:10.1137/1.9781611972788.22.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. (1999). *CRISP-DM 1.0 Step-by-Step Data Mining Guide*.
- Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books.
- Dudani, S.A. (1978). The distance-weighted k-nearest-neighbor rule. *IEEE trans. on systems, man and cybernetics*, 8(4), 311-313.
- Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. John Wiley & Son.
- Liu, C., Cao, L. ve Yu, P. S. (2014). A hybrid coupled k-nearest neighbor algorithm on imbalance data. *International Joint Conference on Neural Networks (IJCNN)*, 2011-2018. doi:10.1109/IJCNN.2014.6889798.
- Liu, W. ve Chawla, S. (2011). Class confidence weighted knn algorithms for imbalanced data sets, In *Pacific-Asia conference on knowledge discovery and data mining*, 345-356. doi:10.1007/978-3-642-20847-8_29.
- Mateos-Garcia, D., Garcia-Gutierrez, J., ve Riquelme-Santos, J.C. (2016). An evolutionary voting for k-nearest neighbors. *Expert Systems with Applications*, 43, 9-14. doi:10.1016/j.eswa.2015.08.017.
- Pang-Ning, T., Steinbach, M., ve Kumar, V. (2006). *Introduction to data mining*, Pearson Education India.
- Parvinnia, E., Sabeti, M., Zolghadri Jahromi, M. ve Boostani, R. (2014). Classification of EEG Signals using adaptive weighted distance nearest neighbor algorithm. *Journal of King Saud University-Computer and Information Sciences*, 26(1), 1-6. doi:10.1016/j.jksuci.2013.01.001.
- Song, Y., Huang, J., Zhou, D., Zha, H. ve Giles, C. L. (2007). IkNN: Informative k-nearest neighbor pattern classification, *European Conference on Principles of Data Mining and Knowledge Discovery*, 248-264. doi:10.1007/978-3-540-74976-9_25.
- Tan, S. (2005). Neighbor-weighted k-nearest neighbor for unbalanced text corpus. *Expert Systems with Applications*, 28(4), 667-671. doi:10.1016/j.eswa.2004.12.023.
- Wang, C., Cao, L., Wang, M., Li, J., Wei, W., ve Ou, Y. (2011). Coupled nominal similarity in unsupervised learning, *Proceedings of the 20th ACM international conference on Information and knowledge management*, 973-978. doi:10.1145/2063576.2063715.
- Yuxuan, L. & Zhang, X. (2011). Improving k nearest neighbor with exemplar generalization for imbalanced classification, *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 321-332. doi:10.1007/978-3-642-20847-8_27.
- Zhang, S., Cheng, D., Deng, Z., Zong, M., ve Deng, X. (2018). A novel kNN algorithm with data-driven k parameter computation. *Pattern Recognition Letters*, 109, 44-54. doi:10.1016/j.patrec.2017.09.036.
- Zhu, Q., Feng, J., ve Huang, J. (2016). Natural neighbor: A selfadaptive neighborhood method without parameter K. *Pattern Recognition Letters*, 80, 30-36. doi:10.1016/j.patrec.2016.05.007.