

Sigorta Hukuk Verisi Üzerinde Karar Ağacı ve Rastgele Orman Algoritmalarının Performans Karşılaştırması

Performance Comparison of Random Forest and Decision Tree Algorithms on Insurance Legal Data

Batuhan BİLENLER¹

Öz:

Bu bilimsel çalışmada, sigorta hukuk datası üzerinde karar ağacı ve rastgele orman sınıflandırıcıları kullanarak belirli parametreler özelinde performans analizi yapılması amaçlanmaktadır. Araştırma için özel bir sigorta şirketinin 15.953 adet verisi kullanılmıştır. Her iki ağaç için dallanma aşamasında gain ratio yöntemi seçilmiştir. Aynı eğitim ve test verileri modeller üzerinde ayrı ayrı uygulanmıştır. Hedef sütunun 1/4 oranında “Evet” olacak şekilde dengesiz dağılım olduğu durumda aşırı ezberlemeyi doğru algoritma ile aşabileceğini göstermiştir. Sınıflandırma aşamasında birden fazla ağaç oluşturan rastgele orman algoritması ile tek bir karar ağacı arasındaki doğruluk oranı, Cohen’s Kappa ve F-measure katsayıları karşılaştırılmış olup; rastgele orman algoritmasının 3 değerlendirme parametresi için de daha başarılı olduğu ortaya koyulmuştur. Model seçiminin anlamlı derecede performans farklılığı ortaya çıkarabileceği sonucuna ulaşılmıştır.

Anahtar Sözcükler: Karar ağacı, rastgele orman algoritması, makine öğrenmesi, veri analizi, veri analitiği

¹ Anadolu Sigorta, İstanbul, Türkiye, E-posta: bilenlerbatuhan@gmail.com

Abstract:

In this scientific study, it is aimed to perform performance analysis on certain parameters by using decision tree and random forest classifiers on insurance law data. For the research, 15.953 data of a private insurance company were used. For both trees, the gain ratio method was chosen at the branching stage. The same training and test data were applied separately on the models. It has shown that if the target column has an unbalanced distribution with a ratio of 1/4 “Yes”, it can overcome overfitting with the correct algorithm. In the classification phase, the accuracy ratio, Cohen's Kappa and F-measure coefficients between the random forest algorithm that generates more than one tree and a single decision tree were compared; It has been revealed that the random forest algorithm is more successful for all 3 evaluation parameters. It has been concluded that the selection of the model may reveal a significant performance difference.

Keywords: Decision tree, random forest algorithm, machine learning, data analysis, data analytics

Giriş

Sigortacılık, finans sektörleri içerisinde veriden para kazanmayı sağlayan en önemli alanlardan biridir. Üretim, hasar, hukuklaşma gibi aşamalarında makine öğrenmesi yöntemleri kullanılabilir. Makine öğrenmesi, geçmiş verileri kullanarak gelecek hakkında çıkarım yapabilmeyi amaçlar. Burada en önemli etken diğer veri madenciliği çalışmalarında olduğu gibi, verinin kalitesi ve erişilebilir olmasıdır (Dolgun et al., 2005). Karlılık, verimlilik, hedef tutturma gibi branşlarda veri kalitesi problemin çözümünde kritik role sahiptir. Veri madenciliği yöntemleri bu gereksinimi karşılamak üzere ortaya çıkmıştır (Özkan, 2008). Veri madenciliği, hem yararlı hem de anlaşılabilir verilerle, alışılmamış yollarla, verileri özetleyen ve gizli ilişkileri ortaya koyan bir analiz yöntemidir (Larose, 2006).

Karar ağaçları, giriş verilerinin bir sınıflandırma veya kümeleme algoritmasıyla, tüm elemanların aynı sınıf etiketine ait olana kadar ağaç dallanmasına benzer şekilde alt gruplara ayırma işlemidir (Orhan, 2012). Sınıflandırma algoritmaları eğitimsiz ve eğitilmiş öğrenme olmak üzere 2 ana gruba ayrılır. Eğitilmiş öğrenme sürecinde veriler önceden etiketlenmiştir. Etiketli veri modellenerek, yeni veriler üzerinde tahminleme yapılır. Test verisi kullanılarak bunların hangi sınıfa dâhil olduğu bulunmaktadır (Aydın, 2007).

Rastgele orman algoritması, ağaç türündeki bir sınıflandırıcıdır. Bu algoritmayı diğerlerinden ayıran en önemli özelliklerden biri rastgele olarak değiştirerek özellik seçimi yapar. Random Forest, tüm değişkenler arasından en iyi dalı kullanarak her bir düğümü dallara ayırmak yerine, her bir düğümde rastgele olarak

seçilen değişkenler arasından en iyisini kullanarak her bir düğümü dallara ayırır (Akar ve Güngör, 2012) Bu strateji Rastgele Orman'nın doğruluğunu eşsiz yapar (Breiman ve Cutler, 2005). Bu yöntem, modelin aşırı öğrenmesini önemli ölçüde engeller.

Bu bilimsel çalışmaya benzer olarak finansal alanlarda makine öğrenmesi algoritmalarının karşılaştırıldığı yayınlara aşağıda yer verilmiştir.

Özgür Erkut Şahin, Serkan Ayvaz ve Engin Çalımfidan kasko sigortası için sahte hasarların tespitinde hasar dosyası incelemelerinde makine öğrenmesi algoritmaları ile çözüm yolları aranmıştır. Veri seti özel bir sigorta şirketinden alınmıştır. Model oluşturulmasında k-NN, regresyon, karar ağacı, YSA gibi yöntemler denenmiş ve sonuçlar irdelenmiştir (Şahin et al., 2020).

Barış Aksoy araştırmasında, sigorta şirketlerinin yılsonu mali tablolarından elde edilen verilerle bir sonraki yıl için Standard & Poors Derecelendirme Kuruluşunun notları baz alınarak Yapay Sinir Ağları, En Yakın k-Komşu Algoritması (KNN) ve Naive Bayes kullanarak tahminleme işlemi yapılmıştır. Sonuç olarak YSA 98.55%, KNN 95.65%, Naive Bayes 85.51% doğruluk oranları elde edilmiştir (Aksoy, 2020)

Oğuz Kaynar , Murat Fatih Tuna, Yasin Görmez ve Mehmet Ali Deveci araştırmasında, makine öğrenmesi algoritmaları kullanarak müşteri kaybı analizi gerçekleştirmiş olup, YSA, Naive Bayes ve SVM algoritmaları arasında performans analizi gerçekleştirmiştir. YSA'nın daha yüksek doğruluk sonucu verdiği ortaya çıkarılmıştır (Kaynar et al., 2017)

Ersin Namlı, Ramazan Ünlü ve Ecem Gül çalışmasında, ikinci el araç satışı tahminlemesi gerçekleştirmiştir. Bu çalışmada makine öğrenme algoritmalarından YSA ve SVM kullanılmış olup, lineer regresyon karşısında ortaya koyacakları performans analiz edilmiştir. Sonuç olarak, makine öğrenmesi algoritmalarının her ikisi de lineer regresyonun ulaştığı sonuçtan daha doğru yaklaşım sergilediği ortaya koyulmuştur (Namlı et al., 2019)

Çalışmada değerlendirilecek parametrelerden ilki doğruluk oranıdır. Her iki model için de sınıflandırma sonuçları içerisinde kaç tanesinin doğru sınıfta tahmin ettiği hesaplanır. Knime Analytics, Scorer işlem nodu ile bu değerleri otomatik olarak hesaplar. Bir modelin başarısı için sadece doğruluk oranına bakılmamalıdır. Veri dağılımının eşit olmadığı durumlarda başka bir değerlendirme parametresi olan F-measure değeri önem kazanır. F1 Score değeri bize Keskinlik (Precision) ve Duyarlılık (Recall) değerlerinin harmonik ortalamasını göstermektedir (Öğündür, 2019). Precision, pozitif belirleme ya da tahminlerin ne kadar doğru olduğunu, Recall ise gerçek pozitiflerin ne kadar doğru belirlendiğini ya da tahmin edildiğini ortaya koyar (Karacı, 2020). Keskinlik ve duyarlılık değerlerinin aritmetik ortalaması alınsaydı, duyarlılık değeri 1 ve keskinlik değeri 0 olan bir senaryoda elde edilen 0.5 F1 score değeri hatalı bir sonuç elde edilmesine yol açacaktı.

Bu araştırmada, sigorta hasar dosyalarındaki bilgiler kullanılarak rastgele orman ve karar ağacı sınıflandırıcıları arasında performans karşılaştırması yapılacaktır.

Rastgele Orman algoritması, hem sınıflandırma hem de regresyon probleminin çözümünde kullanılabilir. Ağaç algoritmalarının en büyük problemlerinden biri aşırı ezberleme olduğu için bu yöntem daha çok tercih edilmektedir. Bu çalışmada sigortacılık verisi üzerinde aşırı ezberlemeye maruz kalan bir model ile ezberlemesi daha az olan model için 3 kriter özelinde bariz performans farkı olması beklenmektedir. Araştırma aynı zamanda model seçiminin performansa etkisinin ne kadar önemli olduğunu da göstermiş olacaktır.

Rastgele Orman algoritması, aşırı ezberlemeyi onlarca yüzlerce ağaç oluşturarak engellemektedir. Yapılmak istenen tahmin veya sınıflandırma işlemleri için farklı yaklaşım sergilemektedir. Günün sonunda problemimiz regresyonsa karar ağaçlarının tahminlerinin ortalamasını problemimiz sınıflandırmaysa tahminler arasında en çok oy alanı seçiyoruz (Şimşek, 2018). Bu algoritma bize seçilen özneliklerin bağımlı değişken üzerinde etkisinin görülmesini de sağlamaktadır. Rastgele Orman algoritması, birden fazla karar ağacı ürettiği için "orman" olarak adlandırılır (Sevgen ve Tanrıvermiş, 2020). Rastgele Orman algoritması, yöntemi hem sosyal bilimler hem de fen bilimleri alanlarında yapılan çalışmalarda modelleme tekniği olarak tercih edilmektedir (Özdemir, 2018).

Materyal ve Metot

Bu araştırmada, sigorta hasar dosyalarında iş birimlerinin yapamayacağı kadar karmaşık problemlerin farklı makine öğrenmesi algoritmalarıyla çözülmesi ve birbirine olan üstün yönlerinin incelenmesi amaçlanmıştır. Veri analizi için, Knime Analytics 4.1.4 kullanılmıştır. Knime yapılması gereken çoğu işlemi hazır “nod” adı verilen işlem kutuları sayesinde kolaylıkla yapabilmektedir. Çalışma için 15.953 veri kullanılmıştır. Performans karşılaştırılması için kullanılan değişkenler ve veri türleri Şekil 1. 'de gösterilmektedir.

Columns: 22	Column Type
DOSYA_RUCU_DURUM	String
ALT_STATU_RUCU_FLG	Number (double)
HUKUK_DOSYASI_ILISKISI	String
MUSTERI_TIPI	String
DOSYA_STATUSU	String
HASAR_SEBEBI	String
RUCU_DURUM	String
TAHMINI_HASAR_TUTARI	Number (double)
HASAR_SAHIBI_ADET	Number (double)
TAM_HASAR_MI	String
FATURALI_IS_MI	String
AS_SKORU	Number (double)
TAZMINAT	Number (double)
HASAR_SAYI	Number (double)
HASAR_KAZA_SAYI	Number (double)
HASAR_RED_SAYI	Number (double)
PLAKA_IL_ADI	Number (integer)
ARAC_YASI	Number (double)
MARKA	Number (integer)
KULLANIM_SEKLI	String
TSB_ARAC_DEGERI	Number (double)
date&time diff	Number (integer)

Şekil 1. Veri setindeki değişkenler ve veri türlerinin gösterimi

Veri seti detaylı şekilde incelendiğinde; HASAR_RED_SAYI, HASAR_KAZA_SAYI ve HASAR_SAYI sütunlarında bazı verilerin eksik geldiği görülmüştür. Bu eksik veriler manipülasyona uğratarak yerine 0 değeri atanacak ve tam veri ile çalışılacaktır.

Araştırmada 21 bağımsız değişken ve 1 bağımlı değişkenden oluşan 15.953 adet veri kullanılmıştır. Veriler öncelikle ön işleme tabi tutulmuştur ve sonrasında modele girdi olarak verilmiştir. Verinin manipülasyona uğratılmış hali Tablo 1’de gösterilmiştir.

Manipülasyon için uygulanan ön işleme algoritmaları aşağıdaki şekilde gerçekleşmiştir:

Hasar Tarihi ile Hasar Açılış Tarihi arasındaki gün farkı arttıkça suistimal şüphesinin arttığı bilindiği için gün sayıları Aşama 1’deki şekilde derecelendirilmiştir.

Aşama 1: Hasar Tarihi ile Hasar Açılış Tarihi arasındaki gün farkını bul. fark ≤ 3 ise 1, $3 < \text{fark} \leq 5$ ise 2, $5 < \text{fark} \leq 7$ ise 3 değilse 4 olarak atama yap.

Veri, Data Explorer nodu yardımıyla incelendiğinde Tazminat miktarı arttıkça hukuklaşma arasında bir ilişki olduğu görülmüş ve Aşama 2’deki gibi veri ön işleme tabi tutulmuştur.

Aşama 2: Tazminat miktarı > Ortalama Tazminat ise 1, değilse 0 olarak atama yap.

Aşama 3: Hedef değişkenin en çok “Evet” olarak geldiği şehirler incelendiğinde en çok “Evet” olan 3 şehir: İstanbul, Ankara, İzmir ise 1; değilse 0 ataması yap.

Aşama 4: Araç bedeli ile hukuklaşma arasında bir ilişki olduğu için, Lüks araç grubu olan BMW, "MERCEDES", "VOLVO-TR", "VOLKSWAGEN", "AUDI", "PORSCH", "RANGE ROVER" ise 1; değilse 0 ataması yap.

Aşama 5: Hasar dosyası üzerinden yapılan ödemelerde anomali durumları önlemek amacıyla: TOPLAM_ODEME_TUTAR_TL 200.000’den büyükse 200.000, -100.000’den küçükse -100.000 değilse kendisini yaz.

Aşama 6: Toplam yapılan ödeme tutarı ile hukuklaşma arasında bir ilişki olduğundan, ortalama toplam ödemenin üzerinde olan değerlerde “Evet” değeri daha çok görülecektir. Bu yüzden, TOPLAM_ODEME_TUTAR_TL > Ortalama TOPLAM_ODEME_TUTAR_TL ise 1; değilse 0 ataması yap.

Aşama 7: Verinin 61%’inde hasar sebebinin çarpma-çarpışma olduğu ve diğer hasar sebeplerine oranla daha fazla hukuklaştığı görülmektedir. Bu sebepten dolayı, HASAR_SEBEBI= “Çarpma” veya HASAR_SEBEBI= “Çarpışma” ise 1 değilse 0 ataması yap.

Aşama 8: Kolonlarda Evet/Hayır olan tüm string değerleri 1/0 olarak değiştirip, üzerine yaz.

Aşama 9: Verinin 55%'ini içerirken, target değerinin 67%'si "Evet" target değerini içerdiği için, KULLANIM_SEKLI = "HUSUSİ OTOMOBİL" veya "KAMYONET-PANEL/GLASS VAN TİPİ" ise 1 değilse 0 olarak atama yap.

Aşama 10: TAHMINI_HASAR_TUTARI'nı Auto-Binner kullanarak 5 farklı kategoriye ayırıp, otomatik olarak muallak tutarları uygun düştüğü aralığa göre 1/2/3/4/5 olarak veriyi dönüştür.

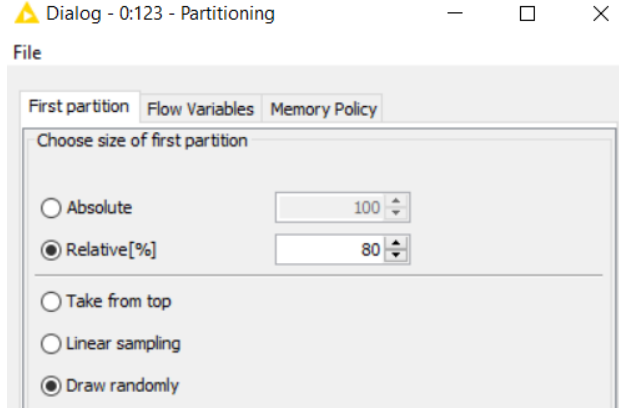
Aşama 11: En son kontrol amacıyla number veya double türünde olup, boş veri içeren saha varsa 0 değerini atama yap.

Yapılan algoritmik ve manipülasyonları sonra son durum Tablo 1'de gösterilmektedir.

Tablo 1. Veri ön işleme sonrası tüm sütunların gösterimi

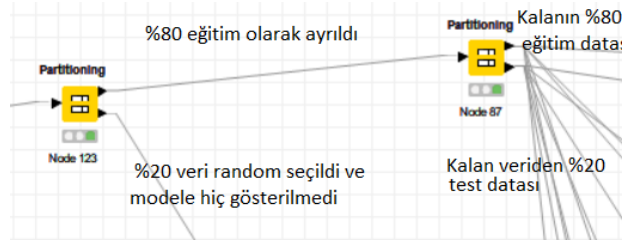
<i>Kolon İsmi</i>	<i>Manipülasyon Öncesi Veri Türü</i>	<i>Manipülasyon Sonrası Veri</i>
<i>DOSYA_RUCU_DURUM</i>	<i>String</i>	<i>1/0</i>
<i>MUSTERI_TIPI</i>	<i>String</i>	<i>1/0</i>
<i>DOSYA_STATUSU</i>	<i>String</i>	<i>1/0</i>
<i>RUCU_DURUM</i>	<i>String</i>	<i>1/0</i>
<i>TAM_HASAR_MI</i>	<i>String</i>	<i>1/0</i>
<i>FATURALI_IS_MI</i>	<i>String</i>	<i>1/0</i>
<i>HASAR_SEBEBI</i>	<i>String</i>	<i>Çarpma:1 Çarpışma:1 Diğerleri :0</i>
<i>KULLANIM_SEKLI</i>	<i>String(Kategorik Değişken)</i>	<i>1/2/3</i>
<i>date&time_diff</i>	<i>Number</i>	<i>1/2/3/4</i>
<i>TAZMINAT</i>	<i>Number</i>	<i>1/0</i>
<i>MARKA</i>	<i>String(Kategorik Değişken)</i>	<i>1/0</i>
<i>PLAKA_IL_ADI</i>	<i>String(Kategorik Değişken)</i>	<i>1/0</i>
<i>TOPLAM_ODEME_TUTAR_TL</i>	<i>Number</i>	<i>1/0</i>

Bu manipülasyon algoritmalarından sonra, verinin 20%'si modele hiç gösterilmeyecek biçimde Partitioning nodu yardımıyla modelden ayrı tutulmuştur. Şekil 2.'de random şekilde ayrılması gösterilmektedir.



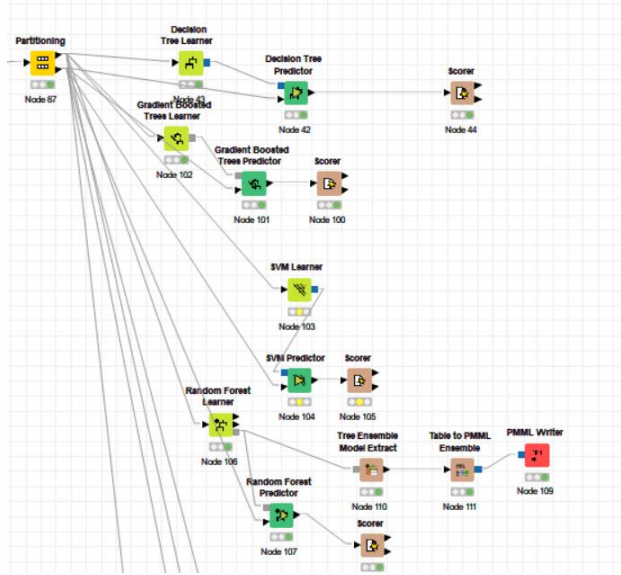
Şekil 2. Rastgele 20% verinin modelden uzaklaştırılması

Geriye kalan verinin 80%'i içinde 20%'lik ikinci bir test verisi de ayrılmıştır. Bu ikinci 20% oranındaki veri modellerin kıyaslanması aşamasında test verisi olarak kullanılacaktır. Şekil 3'te iki kez verinin ayrıldığı gösterilmektedir. Sisteme hiç gösterilmeyen 20% oranındaki data, oluşturulan .pmml modelden okunurken 2. kez test için kullanılacaktır.



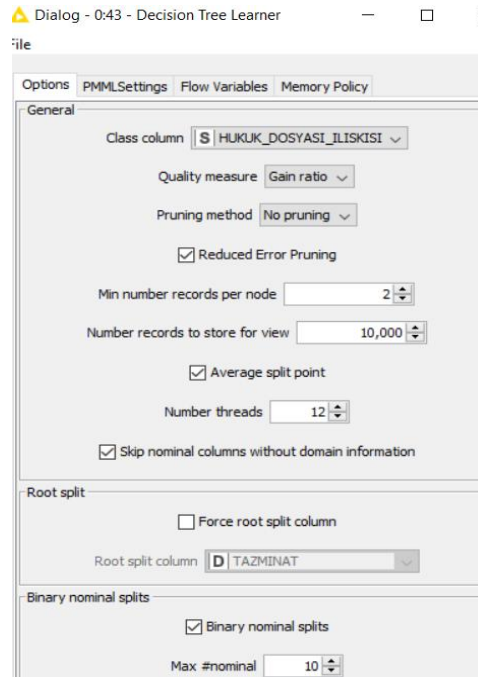
Şekil 3. Verinin random şekilde iki kez parçalara ayrılması

İkinci kez 80% eğitim verisi olarak ayrılan subset Decision Tree Learner ve Random Forest Learner nodlarına girdi olarak verilmiştir. 20% olarak ayrılan test datası ise Decision Tree Predictor ve Random Forest Predictor nodlarına gönderilmiştir. Hukuk dosya ilişkisi, bağımlı değişken 2 farklı yöntem ile sınıflandırılacaktır. Şekil 4'te nodların bağlantıları gösterilmektedir.



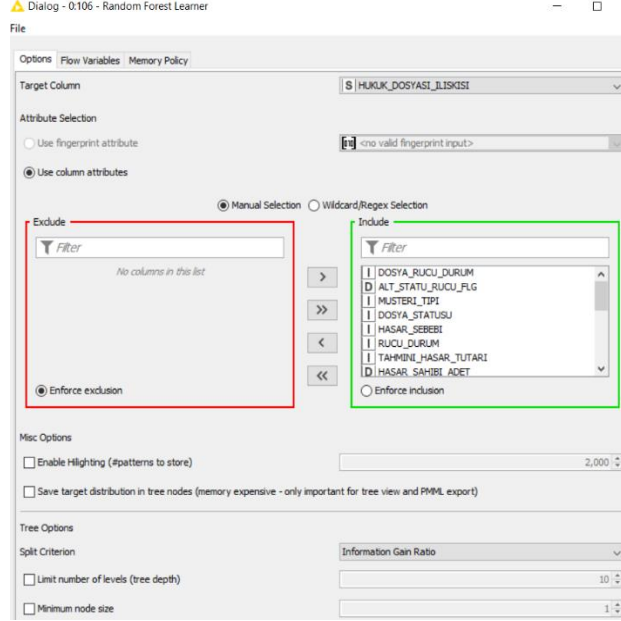
Şekil 4. Sınıflandırıcıların veri ile entegrasyonun gösterimi

Decision Tree modeli oluşturulurken bağımlı değişken olan HUKUK_DOSYASI_ILISKISI kullanılacaktır. Ağaç dallanırken Gain ratio değerine göre oluşturulup, binary olarak ağacın inşa edilmesi sağlanacaktır. Şekil 5'te ağacın özellikleri gösterilmektedir.



Şekil 5. Karar ağacının özellik atamalarının seçilmesi

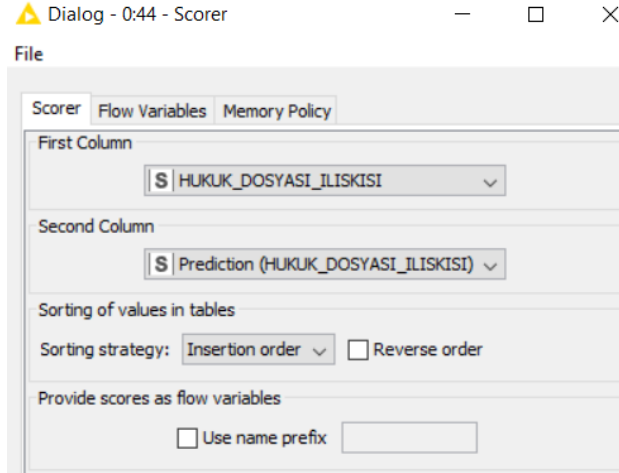
Random Forest sınıflandırıcı için ise, Split kriteri Information Gain Ratio atanmıştır. (Decision Tree ile aynıysa olacak şekilde) Rastsal olarak oluşturulacak ağaçların derinliği max. 10 olması sağlanmıştır. Şekil 6'da sınıflandırıcının özellik atamaları gösterilmektedir.



Şekil 6. Rastgele Orman sınıflandırıcısının özellik atamalarının seçilmesi

Bulgular ve Tartışma

İki sınıflandırıcı model oluşturulduktan sonra Knime üzerindeki Scorer nodu kullanılarak sonuçların analizi yapılacaktır. Şekil 7'de Scorer nodunun yapısı gösterilmiştir. Veri setindeki target ile tahmin edilen değerler kontrol edilip, doğruluk oranının hesaplanması sağlanacaktır.



Şekil 7. Scorer nodunun yapısının gösterimi

Öncelikle karar ağacı üzerinde elde edilen sonuçlar Şekil 8 'de hesaplanmıştır.

D Accuracy	D F-measure	D Cohen's kappa
?	0.924	?
?	0.763	?
0.885	?	0.687

Şekil 8. Karar Ağacı sınıflandırıcısının performans sonuçları

Random Forest sınıflandırıcısı kullanıldığında elde edilen sonuçlar aşağıda gösterilmektedir.

D Accuracy	D F-measure	D Cohen's kappa
?	0.949	?
?	0.829	?
0.922	?	0.778

Şekil 9. Random Forest sınıflandırıcısının performans sonuçları

Confusion Matris değerleri karşılaştırıldığında ise, karar ağacı algoritmasının matrisi Şekil 10'da gösterilmektedir.

Row ID	Hayir	Evet
Hayir	2236	192
Evet	174	589

Şekil 10. Decision Tree Confusion Matrisi

Rastgele Orman algoritmasının matrisi Şekil 11’de gösterilmektedir.

Row ID	Hayir	Evet
Hayir	2339	89
Evet	160	603

Şekil 11. Random Forest Confusion Matrisi

Sonuç

Bu çalışma kapsamında iki farklı yaklaşım ile modeller inşa edilmiş olup, aşağıdaki sonuçlar elde edilmiştir.

- 3 kriter (Accuracy, F-measure, Cohen’s kappa katsayıları) gözetilerek yapılan analizde, her 3 kriter için de Rastgele Orman algoritmasının daha başarılı sonuçlar ortaya koyduğu görülmüştür.
- Confusion matrisler incelendiğinde Rastgele Orman algoritmasının performansının yine daha yüksek olduğu gözlemlenmiştir. İki ağaç için de gain ratio yaklaşımı belirlenmesine rağmen arada anlamlı derecede performans farklılıkları meydana gelmiştir. Birden fazla ağaç oluşturup ezberlemenin nispi olarak engellenmesi modelin daha kararlı yapıda olmasına katkı sağlamıştır. Bu çalışma ayrıca, veriler için doğru modelin seçilmesinin problemlerin çözümüne daha fazla katkı sağladığı sonucunu ortaya koymuştur.
- Veri dengesizliğinin olduğu veri setleri için F-measure değeri göz ardı edilmemelidir.
- Karşılaştırma yapılacak modeller için aynı eğitim ve test kümesinin kullanılması önemlidir.
- Farklı ağaç algoritmaları ile çalışmalar yapıp, daha detaylı araştırma sonuçları elde edilebilir.
- Model oluşturmadan önce veri analizi aşamasında dengeli dağılım, kayıp veri olup olmadığı gibi parametreler incelenmeli ve ona uygun model seçimi yapılmalıdır.

Kaynakça

Dolgun Ö., Balkan C., ve Koç A. A., 2015. Sigortacılık Sektöründe Araç Sigortalarında Suistimal Tespit Sistemi. 2. Ulusal Sigorta ve Aktüerya Kongresi, Karabük, 28-29 Eylül.

Orhan U., 2012. Makine Öğrenmesi, <http://bmb.cu.edu.tr/uorhan/DersNotu/Ders03.pdf>, 06.12.2020.

Aydın S., 2007. Veri madenciliği ve Anadolu Üniversitesi uzaktan eğitim sisteminde bir uygulama, Doktora Tezi, Anadolu Üniversitesi Sosyal Bilimler Enstitüsü.

Özkan Y., 2008. Veri Madenciliği Yöntemleri, Papatya Yayınları.

Larose D. T., 2006. Data Mining Methods and Models, A John Wiley & Sons, Inc., Publication, New Jersey.

Özdemir S., 2018. Random Forest Yöntemi kullanılarak potansiyel dağılım modellemesi ve haritalaması: Yukarıgökdere Yöresi örneği, Türkiye Ormancılık Dergisi, 19(1), sf: 51-56.

Şimşek H. K., 2018. Makine Öğrenmesi Dersleri 5a: Random Forest (Sınıflandırma), <https://medium.com/data-science-tr/makine-%C3%B6%C4%9Frenmesi-dersleri-5-bagging-ve-random-forest-2f803cf21e07>, Erişim Tarihi: 05.12.2020

Sevgen S. C., ve Tanrıvermiş Y. A., 2020. Mass Appraisal With A Machine Learning Algorithm: Random Forest Regression Bilişim Teknolojileri Dergisi, cilt 13, sayı 3, sf: 301-311.,

Şahin E. R., Ayvaz S., ve Çalınfıdan E., 2020. Sigorta Sektöründe Sahte Hasarların Tahmini İçin Geliştirilen Makine Öğrenmesi Modellerinin Kıyaslanması, Bilişim Teknolojileri Dergisi, cilt 13, sayı 4, sf: 479-489.

Aksoy B., 2020. Sigorta Şirketlerinin Derecelendirilmesinde Makine Öğrenmesi Yöntemleri Tahmin Performansının Karşılaştırılması: Türkiye Örneği, Akademik Araştırmalar ve Çalışmalar Dergisi (AKAD), cilt 12, sayı 23, sf: 579-597.

Kaynar O., Tuna M. F., Görmez Y., ve Deveci M. A., 2017. Makine öğrenmesi yöntemleriyle müşteri kaybı analizi, Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi, cilt:18, sayı:1, sf:1-14

Namlı E., Ünlü R., ve Gül E., 2019. Fiyat Tahminlemede Makine Öğrenmesi Teknikleri ve Doğrusal Regresyon Yöntemlerinin Kıyaslanması; Türkiye’de Satılan İkinci El Araç Fiyatlarının Tahminlenmesine Yönelik Bir Vaka Çalışması, Konya Mühendislik Bilimleri Dergisi, cilt 7, sayı 4, sf: 806-821.

Akar Ö., ve Güngör O. 2012. Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması, Jeodezi ve Jeoinformasyon Dergisi, cilt 1 sayı 2, sf: 139-146

Breiman L., and Cutler A., 2005. Random forest, http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm, Erişim Tarihi: 04.12.2020

Öğündür G. 2019., Doğruluk (Accuracy), Kesinlik (Precision), Duyarlılık (Recall) yada F1 Score?, <https://medium.com/@gulcanogundur/do%C4%9Fruluk-accuracy-kesinlik-precision-duyarlı%C4%B1%C4%B1k-recall-ya-da-f1-score-300c925feb38>, Erişim Tarihi: 05.10.2020

Karacı A. 2020, Fiziksel ve Motor Engelli Çocukların Öz Bakım Problemlerinin Derin Sinir Ağları ile Sınıflandırılması, Politeknik Dergisi, 23(2), sf: 333-341