

Finans Alanında Karar Ağacı Sınıflandırıcılarının Kullanımına İlişkin Bir Çalışma

A Study on The Use of Decision Tree Classifiers in The Financial Sector

Batuhan BİLENLER¹

Öz:

Bu bilimsel çalışmada sigorta hasar dosyalarının analiz edilerek, dosyanın mahkemeye gidip gitmeyeceği tahmininin yüksek doğruluk oranıyla yapılması amaçlanmıştır. Araştırma için 27.776 veri kullanılmıştır. Hedef verisinin yarısının “Evet” yarısının “Hayır” olduğu dengeli bir veri setinde karar ağacı yöntemi kullanılarak sınıflandırma işlemi yapılmıştır. Çalışma sonunda, %93.5 doğruluk oranına ulaşılmıştır. Cohen’s Kappa katsayısının 0.86 düzeyinde olması çalışmanın kalitesini ortaya koymuştur. Keskinlik, duyarlılık, F-measure katsayısı muhakkak göz önünde bulundurulmalıdır. Bu çalışmada modelin duyarlılığı hem evet yanıtları için hem de hayır yanıtları için oldukça başarılı sonuç ortaya koymuş olup, her ikisi için de 93% değerinin üzerinde olduğu sonucuna varılmıştır. Sigorta şirketleri için bu kararın önceden tahmin edilebilmesi avukatlık masrafları, dava giderleri, mahkemeye gitmeden daha az tutarla uzlaşma gibi avantajlar sağlayarak yıl sonu bilançolarına büyük destek sağlayacaktır. Bu tür analitik çalışmaların finans alanına ilerde daha fazla katkı sağlayacağı öngörülmektedir.

Anahtar Sözcükler: Veri Analitiği, Karar Ağaçları, Makine Öğrenmesi, Sigortacılık.

JEL Sınıflandırması: C8, O33, G22.

Abstract:

In this scientific study, it is aimed to make the prediction of whether the file will go to court by analyzing the insurance claim files with high accuracy. 27,776 data were used for the research. Classification process was made using the decision tree method in a balanced data set where half of the target data is "Yes" and half is "No". At the end of the study, an accuracy rate of 93.5% was achieved. The fact that Cohen's Kappa coefficient was 0.86 revealed the quality of the study. Precision, sensitivity, F-measure coefficient should definitely be considered. In this study, the sensitivity of the model was quite successful for both yes and no answers, and it was concluded that it was above 93% for both. For insurance companies, the predictability of this decision will

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

provide great support to the year-end balance sheets by providing advantages such as attorney fees, litigation expenses, and settlement with less amount without going to court. It is predicted that such analytical studies will contribute more to the field of finance in the future.

Keywords: Data Analytics, Decision Trees, Machine Learning, Insurance.

JEL Classification: C8, O33, G22.

Giriş

Gelişen teknolojik gelişmeler finansal şirketler için birer fırsat niteliğindedir. Tüm alanlarda köklü değişimlerin olduğu ve bunun da temelinde çok hızlı teknolojik gelişmelerin bulunduğu bir çağda yaşamaktayız (Alkan S.) Veri madenciliği, makine öğrenmesi ve şekil tanıma gibi veri bilimi dallarında her geçen gün içerisinde mevcut öğrenme metotlarının iyileştirilmiş bir versiyonları ve bunun yanında yeni yeni teknikler literatürde yerini almaktadır.(Bulut, 2019) Veriden değer yaratma sanatı olan veri analitiği, doğru kullanıldığı takdirde sektörde fark yaratacak atılımların yapılmasını sağlar. Gerçekten de bilgi teknolojileri, ülke ekonomilerine, verimlilik anlayışına ve örgütsel hayata önemli etkilerde bulunmuş ve pek çok değişimi de beraberinde getirmiştir. (Öktem ve Aydın, 2005) Veri analitiği, sigortacılık sektöründe risk analizi, fiyatlama, kampanya yönetimi, suistimal tespiti, hasar tespiti gibi alanlarda yaygın olarak kullanıldığı bilinmektedir. Değişen teknoloji ve yeniliklerin etkisiyle, günümüz toplumunda insanoğlu bu değişime uyum gösteren çeşitli olaylarla karşılaşmakta ve karşılaştığı olaylardaki riskleri en aza indirebilmek için sigortacılık faaliyetlerinden faydalanmaktadır. (Armağan, 2000)

Sigorta firmaları ile müşterilerin ilişkisi iki temel ayak üzerine inşa edilmiştir: Poliçe üretilmesi ve poliçe sonrası riskin gerçekleşmesi durumunda hasar tarafıdır. Sigorta şirketlerini müşteriden istediği prim tutarını elde etmesi kadar, hasar sürecinde de ortaya koyduğu yaklaşım net karlılık kavramını doğrudan etkileyen parametrelerdir.

Hasar sürecini organize etmek şirketin kendi görevidir. Şirketlerin hasar sürecinde belirli gider kalemleri bulunmaktadır. Araştırmacı ücreti, eksper ücreti, mahkeme masrafları, avukat tutulması, süreye bağlı olarak faiz ödemeleri bunlardan bazılarıdır. Bu bilimsel çalışmada gider kalemlerinden olan mahkeme masrafları, avukat masrafları ve faiz giderlerini ödemeye gerek kalmadan hasar ödemesinin yapılması veya uzlaşma yoluyla şirketten çıkacak olan paranın azaltılması amaçlanmıştır. Mahkeme sürecinden önce tarafların istemesi durumunda sigorta tahkim komisyonuna başvurmak şeklinde de olabilmektedir. Tek koşul mahkemeye gitmeden veya tüketici hakem heyetine aynı konu ile ilgili başvuru yapılmaması şartı ile geçerlidir. Yargılama giderleri içerisinde, uyuşmazlığın avukat eliyle takip edilmesi halinde

ortaya çıkacak vekâlet ücreti de bulunmaktadır. (Özdamar, 2013) Hasar dosyasının mahkemeye gidip gitmeyeceğini tahmin edip, şirkete bu alanda know-how sağlanacaktır. Geçmişte açılmış olan hasar dosyaları ve onların hukuklaşp hukuklaşmama bilgisi kullanılarak makine öğrenmesi algoritmalarından olan karar ağacı yardımıyla tahminleme işlemi gerçekleştirilecektir. Karar ağacı: kök, ağacın dalları, yaprakları ve düğümlerden oluşmaktadır. Ağaçta karar, gidilen nihai yaprağa göre verilir. (J. R. (John Ross) Quinlan and J. Ross, 1993) Sınıflandırma yapılırken veri seti kullanılarak karar ağacı oluşturulur, veriler teker teker ağaca uygulanır ve tahmin için gerekli olan kurallar ortaya çıkarılır (Silahtaroglu, 2013). Ayrıca veri yapılarını ifade etmek için güçlü bir çözüm yolu ortaya koyar (Nizam ve Akın, 2014)

İş analitiği, geçmiş verileri inceleyerek gelecek hakkında çıkarımda bulunma disiplini. Birçok veri analitiği aracı ücretsiz versiyonları ile her çaptaki firmaya yardımcı olduğu görülmektedir. Finans şirketleri iş analitiği araçları yardımıyla büyük karlılık elde etme imkanına kavuşmuştur. Eksik olduğu yönleri ellerindeki ham veriyi işleyerek ortaya çıkarıp, firmanın gelecekteki yol haritasını çizebilme olanağına erişecektir. Bu da verinin ve veri kalitesinin önemini açık şekilde ortaya çıkaracaktır.

Literatür Araştırması

Sigortacılık sektörü, riski paylaşırma ve riski azaltma fonksiyonları yönünden finansal sistem ve reel ekonomide kilit rol oynayan hem ulusal hem küresel bir sektördür. (Ömürbek N. ve Altın F. G., 2008) İş analitiği ile iş zekasını birbirinden ayıran en önemli fark, iş zekası geçmişini incelerken iş analitiği geçmiş verilerini kullanarak gelecek hakkında bilgi vermeyi hedefler. Büyük veri, işletmelerin geleneksel teknolojilere göre çok daha büyük bir büyüme potansiyeline sahip olmasına rağmen hala çok daha az anlaşıldığı halde, büyük verilerin önemini daha hızlı anlayan şirketler, bu konudaki görüşlerini koruyan rekabetlerine karşı pazarda lider konuma getirebilir. Kuruluşlar bu kavramın önemini hafife almamalıdır (Kubinaa, 2015). Günümüzde sürdürülebilir karlılık, istikrarlı büyüme ve nakit akışı kavramları şirketlerin kısa, orta ve uzun vadede geleceği hakkında doğrudan bilgi verecektir.

Şirketlerin hedeflerini yerine getirmek ve fark yaratabilmek için veri kalitesi, sürdürülebilirlik, veri güvenliği gibi alanlara yatırım yapması gerekmektedir. Ancak şirketler, dijital dönüşümün önündeki en büyük engellerin yatırım maliyetlerinin yüksekliği ve yatırımın geri dönüş belirsizliği olduğunu belirtmektedir (Tüsiad, 2017) Müşterilerden toplanan verilerden anlam çıkarabilmek için genellikle analitik toollar kullanılmaktadır. Bu analitik toolların bazıları ücretli bazıları ise ücretsiz özellikler içerebilmektedir. Veriler çok farklı kaynaklardan toplanabilmektedir. Bu nedenle internete bağlı cihazların kaynaklık ettiği verilerin analizinde veri madenciliği yöntemleri yanında web, metin ve multimedya madenciliği teknikleri kullanılmaktadır (Gürsoy, 2017). Sektörde en çok kullanılanları: SAS,

SPSS, Knime gibi platformlardır. Ayrıca Python ve R programlama dili de veri analizi ve veri bilimi alanında etkin kullanılabilir.

Veri Seti

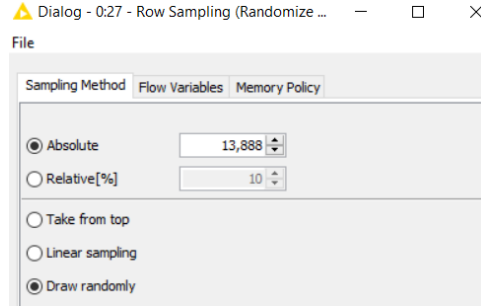
Özel bir sigorta şirketinden alınan veri setinde 26 sütun bilgisi ve bu kolonları kullanarak oluşturulmuş yeni ek bir sütundan oluşmaktadır. Veri analizi için Knime Analytics Platform kullanılmıştır. Bu veri setindeki sütun isimleri ve sütunların veri tipleri Şekil 1’de gösterilmektedir. 27.776 veri bu araştırmada kullanılmıştır. Karar ağaçlarında aşırı ezberleme durumu modelin yanlış kurulmasında etkili faktörlerden biridir. Bu problem bir nebze engelleyebilmek için verilerin seçilirken dengeli dağılım olmasına önem verilmiştir.

Columns: 27	Column Type	Column Index
DOSYA_RUCU_DURUM	String	0
ODEME_STATU_RUCU_F	Number (double)	1
KAYNAK_SISTEM_ID	Number (double)	2
HUKUK_DOSYASI_ILISKISI	String	3
MUSTERI_TIPI	String	4
DOSYA_STATUSU	String	5
ACENTE_IL_ADI	String	6
MADDE_BEDENI_KODU	String	7
RUCU_DURUM	String	8
TAHMINI_HASAR_TUTARI	Number (double)	9
HASAR_SAHIBI_ADET	Number (double)	10
TAM_HASAR_MI	String	11
AS_SKORU	Number (double)	12
ODEME_TUTARI_ORJ_DOV	Number (double)	13
RUCU_TAHSILAT	Number (double)	14
TAZMINAT	Number (double)	15
TOPLAM_ODEME_TUTAR_TL	Number (double)	16
SGR_ODEME_TUTAR_TL	Number (double)	17
EKP_ODEME_TUTAR_TL	Number (double)	18
DGR_ODEME_TUTAR_TL	Number (double)	19
RUC_ODEME_TUTAR_TL	Number (double)	20
HASAR_KAZA_TPL_TUT	Number (double)	21
HASAR_RED_SAYI	Number (double)	22
PLAKA_IL_ADI	String	23
ARAC_YASI	Number (double)	24
MARKA	String	25
date&time diff	Number (integer)	26

Şekil 1. Sütun isimleri ve veri tiplerinin gösterilmesi

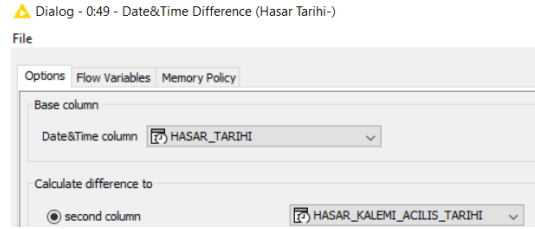
Yöntem

Geçmiş hasar dosyaları bilgilerini içeren 26 sütun karar ağacı algoritmasına input, “HUKUK_DOSYA_ILISKISI” isimli sütun ise output değer olarak verilecektir. Öncelikle veri kümesinde hukuklaşan ve hukuklaşmayan hasar dosyalarının sayıları karşılaştırıldığında hukuklaşan 13.888, hukuklaşmayan 816.301 hasar dosyası olduğu görülmüştür. Karar ağacının daha dengeli ve doğru sonuç vermesi için her iki alt veri kümesinden de eşit sayıda alınacaktır. “Row Sampling” nodu kullanılarak ve “Draw randomly” seçilerek Şekil 2’deki gibi eşit iki alt küme oluşturulmuştur. Bu iki alt veri kümesi concatenate nodu ile alt alta toplanarak yeni veri kümesi daha dengeli hale getirilmiştir.



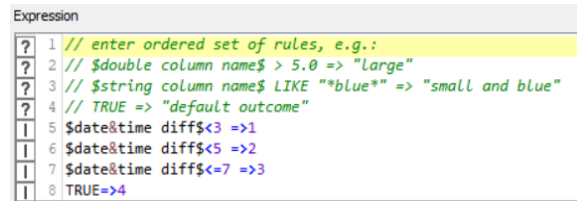
Şekil 2. Rastgele hukuklaşmayan hasar dosyalarının seçimi

Date&Time Difference nodu kullanılarak hasar tarihi ile hasar kalemi açılış tarihleri çıkarılmış ve aradaki gün sayısı karar ağacına **girdi** olarak kullanılmak üzere verilmiştir. Date&Time Difference nodunun kullanımı Şekil 3'te gösterilmektedir.



Şekil 3. Date&Time Difference nodu parameter yapısı ve kullanımı

Elde edilen gün sayısı Rule Engine nodu kullanılarak belirli aralıklar ile gruplandırılmış ve **sayısal** değer almaları sağlanmıştır. Bu kural seti Şekil 4'te gösterilmektedir.



Şekil 4. Gün sayılarının kural setlerine bağlı olarak sınıflandırılması

Veri analizi yapıldığında veri setindeki tazminat miktarı ortalamasının 15.240 TL olduğu görülmüştür. Bu yüzden, 15000 TL'den büyük ve eşit olan değerlerin 1, küçük olanlar için 0 değerleri atanmıştır. Şekil 5'te kural setinin uygulanması gösterilmektedir.

```

Expression
1 // enter ordered set of rules, e.g.:
2 // $double column name$ > 5.0 => "Large"
3 // $string column name$ LIKE "*blue*" => "small and blue"
4 // TRUE => "default outcome"
5 $TAZMINAT$ >= 15000=>1
6 $TAZMINAT$ < 15000=>0

```

Append Column: prediction [I]

Replace Column: [D] TAZMINAT

Şekil 5. Tazminat miktarının manipülasyona uğratılması

```

Expression
1 // enter ordered set of rules, e.g.:
2 // $double column name$ > 5.0 => "Large"
3 // $string column name$ LIKE "*blue*" => "small and blue"
4 // TRUE => "default outcome"
5 $TAHMINI_HASAR_TUTARI$ >= 15000=>1
6 $TAHMINI_HASAR_TUTARI$ < 15000=>0

```

Append Column: prediction [I]

Replace Column: [D] TAHMINI_HASAR_TUTARI

Şekil 6. Tahmini hasar tutarının manipülasyona uğratılması

Verilerin karar ağacına girdi olarak verilmeden önce pre-processing işlemine uğratılması önemli bir adımdır. Yukarıdaki Şekil 6’da tahmini hasar tutarının dönüştürülmesi sağlanmış ve yeni değerler o sütun değerinin üzerine yazılmıştır.

```

Expression
1 // enter ordered set of rules, e.g.:
2 // $double column name$ > 5.0 => "Large"
3 // $string column name$ LIKE "*blue*" => "small and blue"
4 // TRUE => "default outcome"
5 $PLAKA_IL_ADI$ IN ("İSTANBUL", "ANKARA", "İZMİR") =>1
6 TRUE=>0

```

Append Column: prediction [I]

Replace Column: [S] PLAKA_IL_ADI

Şekil 7. Plaka il bilgilerinin transformasyona uğraması

Plaka il adları incelendiğinde ağırlıklı olarak hukuklaşma durumunun belirli şehirlerde kümелendiği gözlemlenmiştir. İstanbul, Ankara ve İzmir şehirleri 1 değeri, diğer 78 şehri içeren hasar dosyaları ise 0 olarak dönüştürülmüştür. Şekil 7’de bu dönüşüm gösterilmektedir. Araçların markaları analiz edilmiş ve lüks araç sınıfına giren bazı araçlarda bağımlı değişkene etkisinin olduğu görülmüştür. BMW,

MERCEDES, VOLVO, VOLKSWAGEN, AUDI, PORSCHE, RANGE ROVER modelleri 1 olarak seçilmiş geri kalan markalar ise 0 olarak Şekil 8 'de gösterilmiştir.

```

Expression
1 // enter ordered set of rules, e.g.:
2 // $double column name$ > 5.0 => "Large"
3 // $string column name$ LIKE "*blue*" => "small and blue"
4 // TRUE => "default outcome"
5 $MARKA$ IN ("BMW", "MERCEDES", "VOLVO-TR", "VOLKSWAGEN", "AUDI", "PORSCHE", "RANGE ROVER") =>1
6 TRUE=>0

```

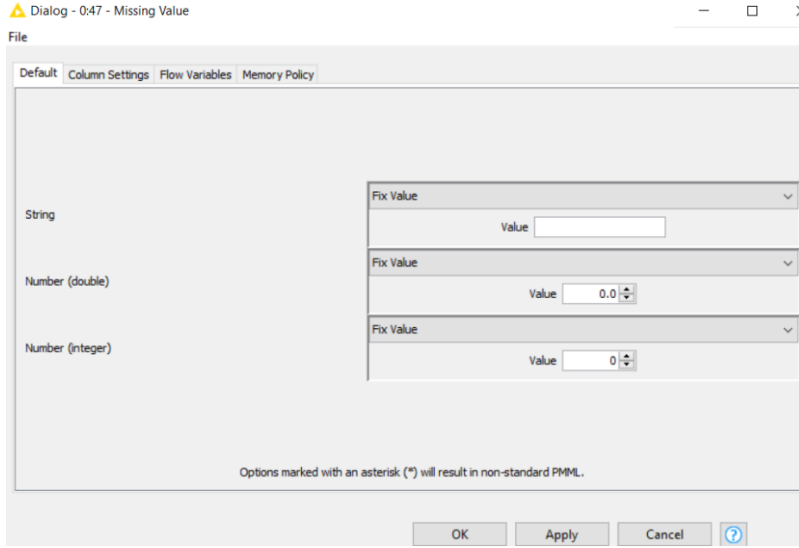
Şekil 8. Marka dönüşümü

Verinin son hali incelendiğinde Şekil 9'da bazı değerlerin boş geldiği gözlemlenmiştir.

U	D	RUCU...	I	TAZMI...	D	TOPLAM_ODEME_TUTAR_TL	D	SGR_ODEME_TUTAR_TL	D	EKP_ODEME_TUTAR_TL	D	DGR_ODEME_TUTAR_TL	D	RUC_ODEME_TUTAR_TL
15.5	15.5	0	?	?	?	?	?	?	?	?	?	?	?	?
138.2	0	0	0	231.24	?	?	?	185.14	886.1	?	?	?	?	-840
3,005.81	1,554.5	0	0	1,170	1,170	?	?	?	?	?	?	?	?	?
-726.38	0	0	0	-148.77	-352.55	?	?	213.78	?	?	?	?	?	-10
-1,141.56	0	0	0	-19,222.71	?	?	?	?	3,642	?	?	?	?	-22,864.71
263.4	0	0	0	308.4	?	?	?	?	308.4	?	?	?	?	?
-522.44	0	0	0	-17,563.38	?	?	?	?	1,548.16	?	?	?	?	-19,111.54
4,442.64	0	0	0	4,470.66	4,210.11	?	?	183.72	76.83	?	?	?	?	?
39,117.5	-36,000	1	39,117.5	74,007.39	1,110.11	?	?	?	?	?	?	?	?	-36,000
3,969.7	0	0	0	4,002.31	3,802.69	?	?	237.38	?	?	?	?	?	-37.76
172.81	0	0	0	260.23	1,639.83	?	?	183.72	768.68	?	?	?	?	-2,332
7,586.61	0	0	0	7,614.63	7,350.46	?	?	183.72	80.45	?	?	?	?	?
-602.98	0	0	0	189.15	-100.26	?	?	183.72	217.2	?	?	?	?	-111.51
126.11	0	0	0	-1,037.77	?	?	?	?	372.23	?	?	?	?	-1,410
2,389.24	0	0	0	2,389.24	2,389.24	?	?	?	?	?	?	?	?	?
1,026.7	0	0	0	?	?	?	?	?	?	?	?	?	?	?
-199.43	0	0	0	?	?	?	?	?	?	?	?	?	?	?
17.35	0	0	0	17.35	17.35	?	?	?	?	?	?	?	?	0
-1,329.96	519.95	0	0	-548.88	?	?	?	?	851.12	?	?	?	?	-1,400
11,532.32	0	0	0	12,077.32	27,218	?	?	359.32	?	?	?	?	?	0
56,794.84	-29,000	1	57,130.43	82,209.38	1,137.98	?	?	2,783.07	?	?	?	?	?	-29,000
3,738	0	0	0	3,740.34	3,625.61	?	?	15.34	99.39	?	?	?	?	?
141,083.57	0	1	141,109.25	140,461.27	168.38	?	?	479.6	?	?	?	?	?	?
129.15	0	0	0	129.15	129.15	?	?	?	?	?	?	?	?	0
1,154.39	0	0	0	1,202.7	1,224.45	?	?	316.71	?	?	?	?	?	-338.46
14,370.69	0	0	0	-7,133.19	-9,164.43	?	?	?	3,779.97	?	?	?	?	-1,686.23
19,534.471	0	1	19,598.55	38,000	420.09	?	?	178.46	?	?	?	?	?	?
1,451.68	0	0	0	1,493.51	1,147.82	?	?	274.27	71.42	?	?	?	?	?
365,361.53	0	1	365,341.43	365,361.53	?	?	?	?	?	?	?	?	?	?
31,433.41	0	1	31,737.16	29,734.81	?	?	?	?	2,002.35	?	?	?	?	?

Şekil 9. Verinin pre-processing işlemleri sonrası son halinin gösterimi

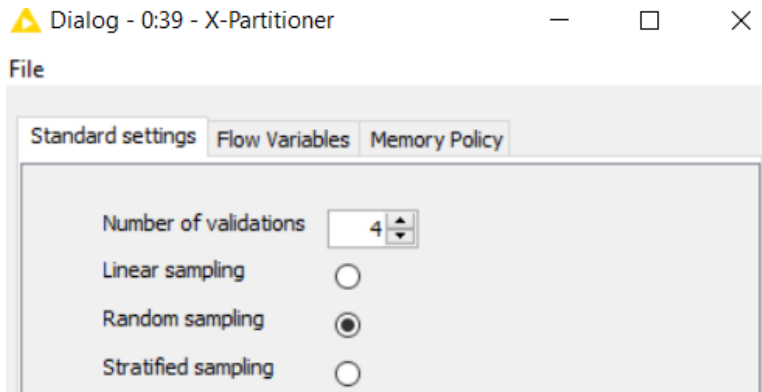
Modelin daha doğru karar verebilmesi için, Missing Value nodu kullanılarak bu verilerin yerine belirli değerlerin atanması sağlanmıştır. String değerler yerine “ ”, double değerler için 0.0, integer değerler yerine 0 ataması gerçekleştirilmiş olup boş verinin olmadığı kontrol edilmiştir. Aşağıda Şekil 10'da missing value nodunun property ayarlarının yapılması gösterilmektedir.



Şekil 10. Missing value nodu kullanımının gösterilmesi

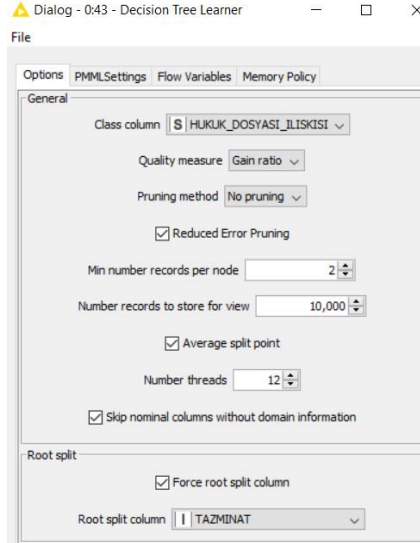
Pre-processing işlemlerinin yapılması ve boş değerlere uygun atamaların gerçekleştirilmesinden sonra 27776 verinin son halde olduğu görülmektedir. Makine öğrenmesi algoritmaları kullanılırken verilerin test ve eğitim olmak üzere 2 alt kümeye ayrılması gerekmektedir. Verilerin %75'i eğitim, %25'i ise test olarak Şekil 11'deki gibi ayrılmıştır. Bu işlem için X-Partitioner nodu kullanılmıştır.

84



Şekil 11. Eğitim ve test verisinin ayrılması

Decision Tree Learner'a verinin %75'i, Decision Tree Predictor noduna %25'i verilmiştir. Decision Tree Learner nodunun özellik atamaları Şekil 12'de gösterildiği şekilde yapılmıştır. Kök düğüm için Tazminat sütunu belirlenmiş olup, hedef değişken olarak "HUKUK_DOSYA_ILISKISI" sütunu modele verilmiştir.



Şekil 12. Model ayarlamalarının gerçekleştirilmesi

Cross Validation işlemi ile verilerin tutarlılığı incelenecek ve rastsallığın ortadan kalkması sağlanacaktır. X-Aggregator nodu kullanılarak oluşturulan 4 alt küme tahminleri toplanıp alt alta tahminlerin yazılması sağlanacaktır. Şekil 13'de makinenin hukuklaşma için tahmin ettiği değerler görülmektedir.

Prediction table - 0:55 - X-Aggregator

File Hilite Navigation View

Table 'default' - Rows: 27776 Spec - Columns: 28 Properties Flow Variables

Row ID	AS_SK...	DOEME...	RUCU...	TAZMI...	TOPLA...	SGR_O...	EXP_O...	DGR_O...	RUC_O...	HASAR...	HASAR...	PLAKA...	ARAC...	MARXA	date&time diff	Prediction (HUKUK_DOSYASI_ILISKISI)
Row194	15	264.86	-3,066.69	0	298.16	3,066.69	218.3	79.86	-3,066.69	0	0	1	1	1	3	Hayir
Row676	66.324	30,140.07	0	1	39,249.02	42,349	982.06	8,760.21	0	55,000	0	0	3	0	1	Evet
Row1078	70	15,941.76	0	0	16,148.2	13,697.26	185.14	2,265.8	0	28,878.67	0	1	3	1	2	Evet
Row1790	66.324	3,603.65	0	0	3,631.67	3,248.48	183.72	199.47	0	3,363,522.24	0	1	1	0	3	Evet
Row2393	81	9,913.66	0	0	9,946.27	9,732.49	213.78	0	0	5,743.7	0	0	2	0	4	Evet
Row2721	56	-121.26	0	0	0	0	0	0	0	0	0	1	13	0	4	Evet
Row2946	26	1,544.22	0	0	1,544.22	1,544.22	0	0	0	336,339.25	0	1	1	0	4	Evet
Row3381	111	1,156.9	0	0	1,185.14	1,000	185.14	0	0	2,500	0	0	7	0	3	Hayir
Row4689	130	95.070	0	1	95,082.6	95,000	82.6	0	0	692,891.64	0	1	6	0	4	Hayir
Row5383	66.324	35,491.98	0	1	35,491.98	35,116.98	0	375	0	0	0	0	10	0	4	Evet
Row5595	36	31,244.85	0	1	31,244.85	31,244.85	0	0	0	736,665.38	2	1	3	0	4	Hayir
Row5806	66.324	3,005.81	1,554.5	0	1,170	1,170	0	0	0	4,769	0	0	2	0	1	Evet
Row6072	62	-522.44	0	0	-17,563.38	0	0	1,548.16	-19,111.54	0	0	1	0	0	1	Evet
Row6073	66.324	4,442.64	0	0	4,470.66	4,210.11	183.72	76.83	0	3,461.88	0	1	15	0	1	Evet

Şekil 13. Hukuklaşma tahmin değerlerinin listelenmesi

Sonuç

Bu bilimsel çalışmanın sonucunda kaç değer doğru tahmin ettiğini görmek için Knime Analytics Platformunun Scorer nodu kullanmış olup sonuçlar Şekil 14’te gösterilmektedir. Accuracy değerinin %93.5 olarak çıkması oldukça yüksek bir oranla doğru tahminleme yapabildiğini göstermiştir.

▲ Accuracy statistics - 0:44 - Scorer

File Hilite Navigation View

Row ID	I TruePo...	I FalsePo...	I TrueNe...	I FalseN...	D Recall	D Precision	D Sensitivity	D Specifity	D F-meas...	D Accuracy	D Cohen'...
Hayır	12939	868	13020	949	0.932	0.937	0.932	0.938	0.934	?	?
Evet	13020	949	12939	868	0.938	0.932	0.938	0.932	0.935	?	?
Overall	?	?	?	?	?	?	?	?	?	0.935	0.869

Şekil 14. Tahmin sonuçlarına ait istatistik bilgilerinin gösterimi

Cohen’s Kappa katsayısının 0.86 çıkması güven aralığının yüksek derece olduğunu ortaya koymaktadır. F-measure değeri kesinlik (precision) ve duyarlılık (recall) değerlerinin harmonik ortalaması alınarak hesaplanan bir metottur. Evet ve Hayır değerleri ayrı ayrı incelendiğinde 0.80 ile 1 değerleri arasında olması karar ağacının güven aralığında sonuçlar ürettiği ortaya koymuştur. Makine öğrenmesi algoritmalarıyla yapılan çalışma sonuçlarında sadece Accuracy değerine odaklanmak yeterli değildir. Keskinlik, duyarlılık, F-measure katsayısı muhakkak göz önünde bulundurulmalıdır. Bu çalışmada modelin duyarlılığı hem evet yanıtları için hem de hayır yanıtları için oldukça başarılı sonuç ortaya koymuş olup, her ikisi için de 93% değerinin üzerinde olduğu sonucuna varılmıştır.

Bu çalışmanın devamı olarak makinenin hukuklaşacak tahmininde bulunduğu dosyaların mahkeme sürecinde kazanılıp kazanılmayacağı üzerine bir araştırma yapılabilir. Bu süreçte de makine öğrenmesi algoritmaları sürecin içinde olacaktır. Veri kalitesi ve veri toplamının önemi dolaylı yoldan ortaya çıkmıştır. Veriden yarar sağlayan ve stratejilerini bu doğrultuda inşa eden şirketlerin ciddi avantajlar yakalayacağı öngörülmektedir. Veriden değer yaratmak isteyen şirketlerin veri analitiğine yatırım yaparak gider kalemlerini düşüreceği, yıl sonu bilançolarına pozitif destek sağlayacakları ortaya koyulmuştur. Bu çalışmaların somut katkıları görüldükçe ileride analitik toolların kullanımının daha da artacağı öngörülmektedir.

Teşekkür

Başta İş Zekası ve Analitik yöneticim Vedat Güneş olmak üzere, değerli arkadaşlarım Hüdayi Özcan, Serkan Kırca, İsmail Çolak'a derin bilgi birikimleri ile makaleye verdikleri fikri destek için teşekkür ederim.

Kaynakça

- Kubinaa, M. Varmusa, M. Kubinova, I. 2015. Use Of Big Data For Competitive Advantage Of Company, 4th World Conference On Business, Economics And Management, WCBEM, Procedia Economics And Finance
- Tüsiad, 2017. Türkiye'nin Sanayide Dijital Dönüşüm Yetkinliği, Aralık 2017 Yayın No: TÜSİAD-T/2017,12 – 589
- Alkan S., 2020. Sigorta Sektörü, <<http://www.vakifbank.com.tr/earastirma/sigorta.doc>>, 10.11.2020.
- Armağan R., 2000, “Yeni Ekonomiler Ve Türkiye”, Süleyman Demirel Üniversitesi İ.İ.B.F. Dergisi, Cilt 5, Sayı 2.
- Öktem M. & Aydın M. 2005, Bilgi Teknolojileri ve Türk Kamu Yönetiminde Dönüşüm, Hacettepe Üniversitesi İ.İ.B.F. Dergisi, Cilt 23, Sayı 2.
- Ömürbek N. & Altın F. G., 2008. Sigortacılık Sektöründe Bilgi Teknolojilerinin Uygulanmasına İlişkin Bir Araştırma Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Sayı 3.
- Özdamar M., 2013. Sigorta Hukukunda Uyuşmazlıkların Çözümünde Tahkim Sistemi. Gazi Üniversitesi Hukuk Fakültesi Dergisi, C. XVII.
- Bulut F., 2019. Bankacılık Sektöründe Yeni Özniteliklerle ve Makine Öğrenmesi Yöntemleriyle Müşteri İlişkileri Yönetiminin Zenginleştirilmesi. Avrupa Bilim ve Teknoloji Dergisi, Sayı 16, S. 382-394
- Şimşek Gürsoy, U. T. 2017. Veri Madenciliğinde Güncel Yaklaşımlar. İstanbul: Çağlayan Kitabevi.

J. R. (John Ross) Quinlan and J. Ross, 1993. C4.5: programs for machine learning,” Morgan Kaufmann Publishers.

Silahtaroglu, G., 2013. Veri Madenciliği Kavram ve Algoritmaları, Papatya Yayıncılık, İstanbul.

Nizam, H. ve Akın, S. S., 2014. Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması, XIX. Türkiye'de İnternet Konferansı