



Çalışanların Çevrimiçi Değerlendirmelerinin İş Tatmini Faktörlerini Belirlemeye ve Analiz Etmeye Yönelik Olarak Makine Öğrenmesine Dayalı Analizi

Ali Özdemir¹, Aytuğ Onan², Vildan Çınarlı Ergene^{3*}

¹Matematik /Fen Edebiyat Fakültesi, Manisa Celal Bayar Üniversitesi, Manisa, Türkiye, (ORCID: 0000 -0001-9330-7084), acaozdemir@gmail.com

² Bilgisayar Mühendisliği / Mühendislik Fakültesi, İzmir Katip Çelebi Üniversitesi, İzmir, Türkiye, (ORCID: 0000-0002-9434-5880), aytugonan@gmail.com

^{3*} Matematik / Fen Bilimleri Enstitüsü, Manisa Celal Bayar Üniversitesi, Manisa, Türkiye (ORCID: 0000-0002-1220-3337), vildan.cinarli@gmail.com

(1st International Conference on Applied Engineering and Natural Sciences ICAENS 2021, November 1-3, 2021)

(DOI:10.31590/ejosat.1008393)

ATIF/REFERENCE: Özdemir, A., Onan, A., & Çınarlı Ergene, V. (2021). İş Tatmini Faktörlerini Belirlemeye Ve Analiz Etmeye Yönelik Olarak Çalışanların Çevrimiçi Değerlendirmelerinin Makine Öğrenmesine Dayalı Analizi. *Avrupa Bilim ve Teknoloji Dergisi*, (28), 526-530.

Öz

Yapay zeka teknolojilerinden biri olan makine öğrenmesi teknikleri günümüzde yaygın olarak kullanılmaktadır. Makine öğrenmesi teknikleri ile hızlı ve güvenilir şekilde tahminlemeler yapılabilmektedir. Makine öğrenmesine ait birçok algoritma mevcut olup, problemin kaynağına ve veri sayısına göre hangi algoritmanın kullanılacağına karar verilmektedir. Farklı algoritmalar, kullanılan veriye göre özgülük ve duyarlılık değerleri üretmektedir. Kullanılan veriye göre bu değerler göz önünde bulundurularak farklı algoritmalar arasından en iyi uyum sağlayan algoritma tercih edilerek sonuç iyileştirilir ve zaman maliyeti azaltılır.

Metin madenciliği son yıllarda hızla gelişen bir alan haline geldi ve bu alandaki uygulamalar yapay zeka alanındaki uygulamaların büyük bir bölümünü oluşturmaktadır. İnsan kaynakları departmanlarının işletme yöneticileri çeşitli konularda kararlar almaktadırlar. Metinsel veriler de dahil olmak üzere daha fazla veri oluşturulup biriktikçe işletme yöneticileri şirketleri ile ilgili karar verirken metin madenciliğini kullanarak bu tür verilerden uygulanabilir bilgi elde edebilirler.

Bu çalışma iş tatmini faktörlerini belirlemeye ve analiz etmeye yönelik olarak çalışanların çevrimiçi değerlendirmelerinin makine öğrenmesine dayalı analizini elde etmek amacıyla gerçekleştirilen çalışmalar hakkında bilgi vermektedir. Çalışanların şirketleri hakkında yaptıkları yorumlardan 1 ve 2 puan olanlar negatif, 3,4 ve 5 puanlar pozitif olarak değerlendirilmiş olup toplamda 18 firmadaki 2321 pozitif yorum ve 702 negatif yorum ele alınarak makine öğrenmesine dayalı analizleri gerçekleştirilmiştir.

Metin belgesindeki özniteliklerin belirlenmesinde terim varlığı, terim sıklığı gibi temel temsil yöntemleri ve 1-gram, 2-gram, 3-gram modelleri dikkate alınarak incelenen veri seti için farklı temsil yöntemlerinin doğru sınıflandırma başarımları, F-ölçütü, duyarlılık, hassasiyet ölçütleri 5 temel sınıflandırıcı aracılığıyla karşılaştırmalı olarak değerlendirilmiştir.

Anahtar Kelimeler: Makine öğrenmesi, Metin madenciliği, Sınıflandırıcı topluluğu

Machine Learning Based Analysis of Employees' Online Assessments to Identify and Analyze Job Satisfaction Factors

Abstract

Machine learning techniques, one of the artificial intelligence technologies, are widely used today. With machine learning techniques, predictions can be made quickly and reliably. There are many algorithms of machine learning, and it is decided which algorithm to use according to the source of the problem and the number of data. Different algorithms produce specificity and sensitivity values according to the data used. Considering these values according to the data used, the best fit algorithm among the different algorithms is preferred, improving the result and reducing the time cost.

Text mining has become a rapidly developing field in recent years and applications in this field constitute a large part of applications in the field of artificial intelligence. Business managers of human resources departments make decisions on various issues. As more data, including textual data, is created and accumulated, business managers can derive actionable insights from such data by using text mining when making decisions about their companies.

This study provides information about the studies carried out to obtain machine learning-based analysis of online evaluations of employees to determine and analyze job satisfaction factors. Among the comments made by the employees about their companies, 1

* Sorumlu Yazar: vildan.cinarli@gmail.com

and 2 points were evaluated as negative, 3,4 and 5 points were evaluated as positive, and a total of 2321 positive comments and 702 negative comments from 18 companies were analyzed, based on machine learning.

Accurate classification performances of different representation methods, F-criterion, sensitivity, precision criteria for the analyzed data set, taking into account the basic representation methods such as term presence, term frequency, and 1-gram, 2-gram, 3-gram models in determining the features in the text document. 5 basic classifiers evaluated comparatively.

Keywords: Machine learning, Text mining, Classifier ensemble

1. Giriş

Her düzeydeki çalışanın bir süre sonunda işinde, çalıştığı kurumda ve iş çevresinde bir dizi deneyimi oluşur. Çalışma yaşamında birtakım beklentileri vardır ve yanı sıra gördükleri, yaşadıkları olaylar ile kazanç ve kayıpları dolayısıyla sevinç ve üzüntüleri olur. Kişi tüm bu deneyimleri sonucu, işine ve iş çevresine karşı tutumlar geliştirir. İş tatmini bireysel-duygusal bir tutum olduğundan açıkça gözlenemez, sadece ifade edilebilir ve ancak bireye sorarak öğrenilebilir. Çalışanın kendisi, işe bakış yöntemiyle işinden ne derece tatmin sağladığını bilebilir. Bir çalışanın işten tatmini işinden ne istediği, kendi geliştirdiği değerlerine ve ekonomik, psikolojik, toplumsal gereksinimlerine dayanır [1].

Günümüzde insanlar işler ve çalıştıkları işyerleri hakkındaki görüşlerini internette ifade edebilmektedirler. Bu kişisel kayıtlar, sosyal psikologlar, pazarlama zekâsı ve fikir madenciliği araştırmaları için zengin ve kullanışlı bir kaynak oluşturmaktadır [2],[3]. Metinsel veriler aynı zamanda büyük ticari değere sahiptir. Şirketler bu verileri müşterilerinin profilini çıkarmaya veya müşteri eğilimlerini anlamaya yardımcı olmak için kullanmaktadır. Bu durum müşterilere daha kişiselleştirilmiş bir deneyim sunmak veya hedefli pazarlama yapmak için bilgi olarak kullanılabilir. Teknolojinin gelişmesi ile çok sayıda çalışan çalışmış oldukları işyerleri hakkında geri bildirimde bulunmaktadır ve araştırmamızda bu değerlendirmeler veri seti olarak kullanılacaktır. Bu verilerden elde edilen sonuçlar, işletme yöneticilerine çalışanlarını yönetme konusunda karar vermelerine ve çalışanlarının iş tatminini çeşitli yönlerden sağlayabilmeleri için faydalı olacaktır.

İş tatmini çalışanın motivasyonu ve performansı ile ilgili olduğu için, firmaların rekabet avantajı elde etmek için çalışanlarının iş tatmini seviyelerinin yüksek olması önemlidir. Konu ile ilgili yapılmış olan araştırmaların çoğu çevrimiçi çalışan incelemelerini analiz eden iş tatmini faktörleri ve firma performansı arasındaki ilişkilere veya iş arayanlar için değer önerilerine odaklanma eğilimindedir. Ancak, çalışanların iş tatmini faktörlerinin çevrimiçi çalışandan belirlenmesi ve bu faktörlere göre yönetsel kararlar elde etme çalışmalarında hala eksiklikler mevcuttur. Bu çevrimiçi değerlendirmeler, çalışanların şirketleri hakkında beğendikleri veya beğenmedikleri yorumları içerir. Yapılan bu değerlendirmeler sayesinde iş tatmini faktörlerini belirlemek ve çalışanların bu faktörler hakkındaki düşüncelerini analiz etmek için çok iyi kaynaklardır.

2. Materyal ve Metod

Literatürde bu alanda birçok çalışma mevcuttur fakat iş tatmini faktörlerine yönelik çalışma pek fazla çalışma bulunmamaktadır. Metin sınıflandırma alanının bir alt dalı olan metinlerde duygu analizi konusunda yaptıkları çalışmada, İngilizce Twitter mesajlarını (tweet), olumlu (positive), tarafsız (neutral) ya da olumsuz (negative) olmak üzere 3 adet sınıf ile sınıflandırmaya çalışmışlardır. Öznitelik vektörü oluşturulurken 1-gram (unigram), 2-gram (bigram) ve POS öznitelik temsili

yöntemleri kullanılırken, sınıflandırıcı olarak NaiveBayes, Maksimum Entropi ve Destek Vektör Makineleri kullanılmıştır [4].

Çalışmamızda çalışanların çevrimiçi değerlendirmelerinin sınıflandırıcı topluluklarına dayalı analizi yapılması amacıyla çalışan kişilerin çalışmış oldukları firmalar hakkındaki düşünceleri tr.indeed.com'da yayınlanan toplam 18 farklı firmadan elde edilen 3023 adet çevrimiçi değerlendirme elde edilmiştir. Sistemde kullanılacak olan veriler ilk olarak WEKA programında hataları düzeltilerek ön işlemden geçirilmiştir. Ön işlemden geçirilen çevrimiçi değerlendirmeler, öznitelik temsili yöntemlerinden terim sıklığı (TermFrequency-TF), terim varlığı (TermPresent-TP) ve TF-IDF ölçütleri için N-gram temsil modellerinden olan 1-gram, 2-gram ve 3-gram temsilleri ile toplam dokuz farklı veri temsili elde edilmiştir. Daha sonra çıkarılan öznitelikler ağırlıklandırılmıştır ve elde edilen verileri arff (Attribute Relational File Format) formatına dönüştürülerek sınıflandırma işlemi gerçekleştirilmiştir. 5 temel sınıflandırıcı olarak NaiveBayesMultinomial, lazy.IBk, Random.forest, functions.Logistic ve functions.LibLINEAR kullanmış ve makine öğrenmesine dayalı deneysel sonuçlar yapılmıştır.

Çevrimiçi değerlendirme yapılan firmaların adı ve kaç adet pozitif ve negatif yorum yapıldığını belirten bilgiler aşağıdaki tablo 1'de verilmiştir:

Tablo 1. Çevrimiçi değerlendirme yapılan firmalara ait yorum sayıları

Firma adı	Pozitif (3-4-5 puan)	Negatif (1-2 puan)	Toplam Firma Sayısı
VESTEL	85	16	101
BOSCH	37	3	40
ARÇELİK	151	23	174
SIEMENS	33	2	35
MC DONALDS	151	54	205
BURGER KING	170	49	219
DOMİNOS	13	21	89
KOTON	0	43	173
LC WAIKIKI	337	49	386
DE FACTO	189	32	221
STARBUCKS	86	8	94
KAHVE DÜNYASI	38	16	54
TURKCELL	0	24	158
TURK TELEKOM	0	33	212
VODAFONE	0	23	161
BİM A.Ş.	113	39	152
A101	248	260	508
MIGROS	34	7	41

2.1. Deneysel Süreç

Terim ağırlıklandırma bir terimin döküman içerisinde ne kadar önemli olduğunu belirtmek amacıyla oluşturulan terim vektöründe her bir terim için ağırlık belirleme işlemidir [4]. Literatürde terim ağırlıklandırma ile ilgili çalışmalar mevcuttur. Kısa mesajları önceden belirlenmiş sınıflara göre sınıflandırma işlemi öncesinde, öznitelik vektörlerinin oluşturulması ve ağırlıklandırma metodları ile başarımın artırılması konusu üzerinde durmuşlardır. Çalışmada yaygın kelimeler (commonwords) değerlendirmeden çıkarıldıktan sonra, kalan kelimelerin kökleri saptanmış ve Open NLP adlı bir araç vasıtasıyla isim, sıfat ve fiiller etiketlenmiştir. Son olarak elde edilen bu vektör uzay modeli TF-IDF yöntemi ile ağırlıklandırılmıştır. Sınıflandırma öncesi gerçekleştirilen bu adımların, başarıyı arttırdığı gözlenmiştir [5].

Yaptığımız çalışmada ise özniteliklerin ağırlıklandırılması için TF (Terim Frekansı), TP (Terim Varlığı) ve TF-IDF (Terim Frekansı-Ters Belge Frekansı) olmak üzere üç farklı ağırlıklandırma yöntemi uygulanarak 1-gram, 2-gram ve 3-gram temsilleri ile toplam dokuz farklı konfigürasyon aşağıdaki şekilde elde edilmiştir:

- IDF FALSE
TF TRUE

OUTPUTWORDCOUNTS TRUE alınarak;
TF_unigram, TF_bigram ve TF_trigram oluşturuldu,

- IDF TRUE
TR TRUE

OUTPUTWORDCOUNTS TRUE alınarak;
TFIDF_unigram, TFIDF_bigram ve TFIDF_trigram oluşturuldu,

- IDF FALSE

TF FALSE alınarak TP_unigram, TP_bigram ve TP_trigram oluşturuldu.

3. Araştırma Sonuçları ve Tartışma

Weka tabanlı Java uygulaması ile elde edilen toplam 9 adet konfigürasyonun analizleri için 5 temel sınıflandırıcı olarak NaiveBayesMultinomial, lazy.IBk, Random.forest, functions.Logistic ve functions.LibLINEAR kullanmıştır. Buna göre toplamda 45 tane deney elde edilmiş oldu. Deneysel analizlerin sonuçları aşağıda Tablo2.'de sunulmuştur.

Tablo2.Temel sınıflandırma algoritmalarına dayalı analizler

Algoritma	Veri seti için elde edilen sonuçlar				
NaiveBayesMultinomial	CorrectlyClassifiedInstances	Precision	Recall	F-Measure	Time takentobuild model:
TF_bigram.arff	80.6484%	0,79	0,806	0,775	0.02 seconds
TF_trigram.arff	78.2005%	0,772	0,782	0,711	0 seconds
TF_unigram.arff	83.0301%	0,82	0,83	0,821	0 seconds
TFIDF_bigram.arff	76.6457%	0,766	0,766	0,766	0 seconds
TFIDF_trigram.arff	73.0731%	0,701	0,701	0,712	0 seconds
TFIDF_unigram.arff	80.4830%	0,807	0,805	0,806	0 seconds
TP_bigram.arff	80.2845%	0,785	0,785	0,786	0 seconds
TP_trigram.arff	76.3480%	0,721	0,763	0,725	0 seconds
TP_unigram.arff	82.1700%	0,814	0,822	0,817	0 seconds
lazy.IBk	CorrectlyClassifiedInstances	Precision	Recall	F-Measure	Time takentobuild model:
TF_bigram.arff	72.2792%	0,693	0,723	0,705	0 seconds
TF_trigram.arff	75.1571%	0,679	0,752	0,69	0 seconds
TF_unigram.arff	68.7727%	0,667	0,688	0,676	0 seconds
TFIDF_bigram.arff	71.9153%	0,69	0,719	0,702	0 seconds
TFIDF_trigram.arff	75.1902%	0,68	0,752	0,69	0 seconds
TFIDF_unigram.arff	70.3936%	0,664	0,704	0,68	0 seconds
TP_bigram.arff	74.4955%	0,698	0,745	0,71	0 seconds
TP_trigram.arff	75.3556%	0,681	0,754	0,69	0 seconds
TP_unigram.arff	70.3936%	0,664	0,704	0,68	0 seconds
Random.forest	CorrectlyClassifiedInstances	Precision	Recall	F-Measure	Time takentobuild model:
TF_bigram.arff	76.3480%	0,734	0,763	0,742	130.81 seconds
TF_trigram.arff	75.9841%	0,704	0,76	0,707	231.94 seconds
TF_unigram.arff	81.4754%	0,806	0,815	0,782	50.23 seconds
TFIDF_bigram.arff	76.5134%	0,735	0,765	0,742	128.17 seconds
TFIDF_trigram.arff	75.7195%	0,7	0,757	0,704	227.46 seconds

Tablo 2'nin devamı

TFIDF_unigram.arff	81.5084%	0,807	0,815	0,782	47.77 seconds
TP_bigram.arff	76.3480%	0,735	0,763	0,742	127.26 seconds
TP_trigram.arff	74.8263%	0.69	0.748	0.701	57.24 seconds
TP_unigram.arff	78.6636%	0.766	0.787	0.77	16.25 seconds
functions.Logistic	Correctly Classified Instances	Precision	Recall	F-Measure	Time taken to build model:
TF_bigram.arff	67.7473%	0,687	0,677	0,682	394.77 seconds
TF_trigram.arff	71.7499%	0.682	0.717	0.696	46.31 seconds
TF_unigram.arff	66.7549%	0.696	0.668	0.68	16.11 seconds
TFIDF_bigram.arff	68.4750%	0.691	0.685	0.688	436.73 seconds
TFIDF_trigram.arff	71.6176%	0.681	0.716	0.695	45.63 seconds
TFIDF_unigram.arff	67.4826%	0.701	0.675	0.686	13.64 seconds
TP_bigram.arff	68.4419%	0.692	0.684	0.688	191.98 seconds
TP_trigram.arff	71.9153%	0.684	0.719	0.697	29.55 seconds
TP_unigram.arff	67.4826%	0.701	0.675	0.686	13.1 seconds
functions.LibLINEAR	Correctly Classified Instances	Precision	Recall	F-Measure	Time taken to build model:
TF_bigram.arff	76.9434%	0,739	0,769	0,745	0.07 seconds
TF_trigram.arff	76.1495%	0,705	0,761	0,704	0.02 seconds
TF_unigram.arff	77.7373%	0,766	0,777	0,771	0.04 seconds
TFIDF_bigram.arff	72.1138%	0,71	0,721	0,715	0.12 seconds
TFIDF_trigram.arff	74.1317%	0,687	0,741	0,7	0.09 seconds
TFIDF_unigram.arff	71.6838%	0,731	0,717	0,723	0.15 seconds
TP_bigram.arff	71.6838%	0,731	0,717	0,723	0.02 seconds
TP_trigram.arff	75.5210%	0,7	0,755	0,707	0.02 seconds
TP_unigram.arff	75.7526%	0,751	0,758	0,754	0.06 seconds

Metin madenciliği yöntemleri, verisetlerindeki önemli bilgileri izole etmeye ve zorlukları fırsatlara dönüştürmeye yardımcı olur. Bu çalışmanın temel amacı metin madenciliğini kullanarak, işyerinde çalışanlara gerekli olan iş tatmini gereksinimlerindeki değişen eğilimleri ve koşulları belirlemektir. Metin madenciliği, metin belgelerinden bilgi veya bilgi çıkarmak için kullanılan genellikle yapılandırılmamış bilgiler içeren bir tekniktir [7].

Yapılandırılmamış bilginin analiz edilmesi zordur çünkü doğrudan analiz edilebilir dijital bilgidir; ancak metin madenciliği yöntemleri bu sorunu çözebilir. Metin madenciliği teknikleri, özellikle yükseköğretim alanında olmak üzere çeşitli alanlarda yaygın olarak benimsenmiştir; ayrıca bu teknik öğrencilerin öğrenme performansını değerlendirmek için birçok çalışmada kullanılmıştır [8]. Bu nedenle, bu çalışmada, insan kaynakları yönetiminde iş tatmin faktörlerini makine öğrenmesine dayalı analiz etmek için metin madenciliği yöntemi kullanılmıştır.

4. Sonuç

İnsan kaynaklarında yapay zeka, bilişsel bilim ve bilişsel davranış modellemesinin anlamasına yardımcı olmakta ve geçmiş, şimdiki ve gelecekteki olayların etkili bir analizini içermektedir. Geçmiş olayların incelenmesi "Raporlama", mevcut olayların incelenmesi "Analiz", gelecekteki olayların veya beklentilerin incelenmesi "Öngörülmesi" olarak bilinmektedir. Yapay zeka araçlarıyla üçünün de birlikte çalışılması, örgütün genel organizasyon hedeflerine ulaşması için

farklı insan kaynakları faaliyetlerinin etkin bir şekilde planlanmasına ve tahmin edilmesine yol açacaktır [9].

İnsan kaynakları yönetiminde makine öğrenmesinin kullanılması; iş için doğru adayları seçmekte ve yetenek kazanmalarına yardımcı olmakta, işyerinde çalışanların elde tutulma oranının tahmin edilmesine ve şirketteki idari personel üzerindeki yükü azaltmaya yardımcı olmaktadır. Aynı zamanda, hata olasılığını minimum seviyeye indirmeye, çalışanların daha yüksek performans ile katılım sağlamasına ve karar vermede önyargı davranışını en aza indirmeye yardımcı olmaktadır [10]. Bu çalışmaların sonucunda iş tatmini faktörleri analiz edilerek çalışanların iş doyumunun artmasına olanak sağlanır.

Tablo 2.'de verilen deney sonuçlarına göre en düşük başarı oranı ise 66.7549% ile functions.Logistic algoritmasında Terim sıklığı (TermFrequency) ve 1-gram modelinde yani TF_unigram.arff konfigürasyonu kullanıldığında elde edilmiştir. Deney sonucu 16.11 saniyede alınmış olup elde edilen duyarlılık değeri 0,696; hassasiyet değeri 0,668 ve F-ölçütü değeri ise 0,68 dir. Yine Tablo 2.'de görülen bu analizler sonucunda en yüksek başarı oranı 83.0301% ile NaiveBayesMultinomial algoritmasında TF_unigram.arff konfigürasyonundan yani Terim sıklığı (TermFrequency) ve 1-gram modeli kullanılarak 0 saniye gibi çok kısa sürede analiz sonucu elde edilmiştir. Bu sonuçlara göre duyarlılık değeri 0,82; hassasiyet değeri 0,83 ve F-ölçütü 0,821 olarak bulunmuştur.

Kaynakça

- [1] Kantar, Hüseyin (2014), İşletmede Motivasyon, Kum Saati Yayınları, İstanbul.
- [2] Bollen, Johan, Huina Mao, and Xiaojun Zeng. "Twitter mood predicts the stock market." *Journal of Computational Science* 2.1 (2011): 1-8.
- [3] Prabowo, Rudy, and Mike Thelwall. "Sentiment analysis: A combine dapproach." *Journal of Informetrics* (2009): 143-157.
- [4] Ozkaya, U., Melgani, F., Bejiga, M. B., Seyfi, L., Donelli, M., GPR B scan image analysis with deep learning methods. *Measurement*, 165, 107770.
- [5] Salton, Gerard, and Christopher Buckley. "Term-weighting approaches in automatic text retrieval." *Information processing and management* 24.5 (1988): 513-523.
- [6] Al-Talib, G. A. and Hassan, H. S., 2013, A study on analysis of SMS classification using TF-IDF Weighting, *International Journal of Computer Networks and Communications Security*,1(5), 189-194pp.
- [7] Ganesh, S. (2013), "Text mining in analyzing the presentation of educational trainers", *International Journal of Computer Applications*, Vol. 66 No. 7, pp. 38-44.
- [8] Abdous, M. And He, W. (2011), "Using text mining to uncover students' technology-related problems in live video streaming", *British Journal of EducationalTechnology*, Vol. 42 No. 1, pp. 40-49.
- [9] Jain, S., (2018), "Human Resource Management and Artificial Intelligence", *International Journal of Management and Social Sciences Research (IJMSSR)*, Volume 7, No. 3, pp.56-59.
- [10] Yawalkar, V. V., (2019), "Study of Artificial Intelligence and its role in Human Resource Management", *International Journal of Research and Analytical Reviews (IJRAR)*, pp.20-24).