İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

TFEEC : TÜRKÇE FİNANSAL OLAY ÇIKARIM DERLEMİ

YÜKSEK LİSANS TEZİ

Kadir Şinas KAYNAK

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

HAZİRAN 2022

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

TFEEC : TÜRKÇE FİNANSAL OLAY ÇIKARIM DERLEMİ

YÜKSEK LİSANS TEZİ

Kadir Şinas KAYNAK
(504191525)

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Doç. Dr. A. Cüneyd TANTUĞ

HAZİRAN 2022

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

TFEEC : TURKISH FINANCIAL EVENT EXTRACTION CORPUS

M.Sc. THESIS

Kadir Şinas KAYNAK
(504191525)

Department of Computer Engineering

Computer Engineering Program

Thesis Advisor: Doç. Dr. A. Cüneyd TANTUĞ

JUNE 2022

İTÜ, Lisansüstü Eğitim Enstitüsü'nün 504191525 numaralı Yüksek Lisans Öğrencisi Kadir Şinas KAYNAK, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı "TFEEC : TÜRKÇE FİNANSAL OLAY ÇIKARIM DERLEMİ" başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Doç. Dr. A. Cüneyd TANTUĞ** .....
İstanbul Teknik Üniversitesi

Jüri Üyeleri : **Prof. Dr. Banu DİRİ** .....
Yıldız Teknik Üniversitesi

Doç. Dr. Yusuf YASLAN .....
İstanbul Teknik Üniversitesi

Teslim Tarihi : **1 Haziran 2022**
Savunma Tarihi : **5 Temmuz 2022**

Aileme ve arkadaşlarıma,

ÖNSÖZ

Yoğun efor ve özveri ile hazırlamış olduğum yüksek lisans tezimi tamamlamanın mutluluğunu ve heyecanını yaşarken, benden yardımlarını ve desteklerini esirgemeyen o güzel insanlara teşekkürü borç bilirim.

Öncelikle yüksek lisans eğitimim boyunca danışmanlığımı üstlenen, araştırma konumun belirlenmesinde ve yürütülmesinde ilgisini ve desteğini esirgemeyen, deneyimiyle yol gösteren hocam Doç. Dr. Ahmet Cüneyd TANTUĞ'a sonsuz teşekkürler.

İTÜ'de bizim gibi mühendislerin yetiştirilmesinde büyük emeği geçen, verdikleri derslerle bizleri donatan tüm İTÜ öğretim görevlilerine de tek tek teşekkürlerimi sunarım.

Ve, desteğini hiç esirgemeyen aileme sevgilerimi sunuyorum.

NİSAN 2022 Kadir Şinas KAYNAK

İÇİNDEKİLER

Sayfa

ÖNSÖZ ..... **ix**
İÇİNDEKİLER ..... **xi**
KISALTMALAR ..... **xiii**
SEMBOLLER ..... **xv**
ÇİZELGE LİSTESİ ..... **xvii**
ŞEKİL LİSTESİ ..... **xix**
ÖZET ..... **xxi**
SUMMARY ..... **xxv**
1. GİRİŞ **1**
1.1 Tezin Amacı 1
1.2 Tezin Katkıları 2
1.3 Genel Bakış 3
2. LİTERATÜR ARAŞTIRMASI **5**
2.1 Olay Çıkarımı Derlemi 5
2.2 Veri Genişletme Metotları 6
2.2.1 Kendi kendine eğitim 6
2.2.2 Aktif öğrenme 6
2.2.3 Öğrenme aktarımı 7
2.2.4 Zayıf denetim 7
3. OLAY ÇIKARIMI **9**
3.1 Olay 9
3.2 Olay Tetikleyicisi 10
3.3 Olay Argümanı 14
4. VERİ KÜMESİ OLUŞTURMA **17**
4.1 İşaretleme Süreci 17
4.2 İşaretleme Aracı 21
5. VERİ KÜMESİ GENİŞLETME YÖNTEMİ **23**
5.1 Aktif Öğrenme 23
5.1.1 Rastgele örneklem seçimi (Random selection) 23
5.1.2 En az güven seçimi (Least confidence selection) 23
5.1.3 En çok güven seçimi (Max confidence selection) 24
5.1.4 Marj belirsizliği seçimi (Margin selection) 24
5.1.5 Entropi seçimi (Entropy selection) 24
5.2 Zayıf Denetleme 25
5.3 Zayıf Denetleme ile Geliştirilmiş Aktif Öğrenme Döngüsü 26
6. DEĞERLENDİRMELER **31**
6.1 Olay ve Argüman Çıkarımı 31
6.2 Veri Kümesi 34
7. SONUÇLAR **37**
KAYNAKLAR **39**
EKLER **43**
EK A : Olay Türü ve Argüman Sınıflandırma Sonuçları 45
ÖZGEÇMİŞ **49**

KISALTMALAR

SVM : Destek Vektör Makinesi (Support Vector Machine)
ME : Maksimum Entropi (Maximum Entropy)
RNN : Tekrarlayan Sinir Ağları (Recurrent Neural Networks)
GNNs : Grafik Sinir Ağları (Graph Neural Networks)
CNN : Konvolüsyonel Sinir Ağları (Convolutional Neural Networks)
NER : Varlık İsmi Tanıma (Named Entity Recognition)
AL : Aktif Öğrenme (Active Learning)
WS : Zayıf Denetim (Weak Supervision)
QS : Sorgu Seçimi (Query Strategy)
EC : Olay Sınıflandırma (Event Classification)
AC : Argüman Sınıflandırma (Argument Classification)

SEMBOLLER

UH	**:** Etiketlenmemiş Veriler (Zor)
UE	**:** Etiketlenmemiş Veriler (Kolay)
T	**:** Olay Etiketi
N	**:** Varlık İsmi Tanımlama
f	**:** Sınıflandırıcı
y	**:** Etiketler

ÇİZELGE LİSTESİ

Sayfa

Çizelge 4.1 : Olay tür ve alt tür listesi. **18**
Çizelge 6.1 : Temel model sonuçları..... **31**
Çizelge 6.2 : Olay sınıflandırma sonuçları. **32**
Çizelge 6.3 : Argüman sınıflandırma sonuçları. **32**
Çizelge 6.4 : Veri kümelerine ait istatistiksel bilgiler. **34**
Çizelge A.1 : Argüman sınıflandırma modelinin sınıf bazındaki sonuçları. **45**
Çizelge A.2 : Olay sınıflandırma modelinin sınıf bazındaki sonuçları. **46**
Çizelge A.3 : Olay sınıflandırma modelinin sınıf bazındaki sonuçları (devam)..... **47**
Çizelge A.4 : Olay sınıflandırma modelinin sınıf bazındaki sonuçları (devam)..... **48**

ŞEKİL LİSTESİ

Sayfa

Şekil 4.1 : İşaretlenmiş örnek bir cümle. **19**
Şekil 4.2 : Türkçe finansal olay veri kümesindeki olay dağılımı. **20**
Şekil 4.3 : WebAnno işaretleme platformu. **22**
Şekil 5.1 : Zayıf denetim sürecinin mimarisi. **26**
Şekil 5.2 : Tasarlanmış modelin mimarisi. **27**
Şekil 5.3 : Zayıf denetim modülü. **27**

TFEEC : TÜRKÇE FİNANSAL OLAY ÇIKARIM DERLEMİ

ÖZET

Gelişen dünya ile dijitalleşmeye olan ilgi gitgide artmaktadır. Son zamanlarda çeşitli alanlarda yapılan dijitale dönüşüm çalışmaları da bu trendin önemini vurgulayan işler oldu. Dijitalleşme sayesinde inanılmaz sayılarda metinler üretilmekte ve paylaşılmaktadır. Bu dijital kaynaklardan bilginin çıkarımı az sayıdaki veri için elle yapılabilir olsa da çok büyük miktarlardaki verilerle çalışmak oldukça yoğun ve zaman alıcı olmaktadır. Yaşanan problemleri aşabilmek ve elle yapılan çalışmaları otomatikleştirebilmek için bilgi çıkarma teknikleri geliştirilmeye başlanmıştır. Bilgi çıkarma, genellikle metinlerden doğal dil işleme teknikleri kullanılarak yapılandırılmış bilgilerin otomatik olarak çıkartılmasını hedefler. Bu da bu alandaki insan eforunu ve masrafları azaltarak süreci daha verimli hale getirir. Şu ana kadar yapılan katkılar ve gelecekte sağlayacağı düşünülen faydalardan dolayı bu alanda yapılan çalışmalar popülerlik kazanmıştır.

Bilgi çıkarmanın bir türü de olay çıkarımıdır. Olay çıkarma metindeki olayları bulmayı, bulunan olayların tiplerinin tespitinin yapılmasını ve buna karşılık gelen argümanların tanımlanmasını içeren zorlu bir görevdir. Bu sayede elde edilen yapılandırılmış bilgiler, farklı doğal dil işleme görevlerine (bilgi tabanı oluşturma, soru cevaplama, dil anlama gibi) faydalı bilgiler sağlayabilir ve katkıda bulunabilir. Ancak bu alanda sınırlı sayıda literatür çalışması yapılmıştır ve yapılan çalışmalar da genellikle belirli bir alana odaklanır. Küresel çapta yapılan olay çıkarımı araştırmalarının odaklandığı çalışma alanlarından biri de finans ve ekonomidir.

Finansal olay çıkarımı hem şirketler hem de yatırımcılar için çok önemlidir. Şirketler ürünleri hakkında hızlı geri dönüş alabilmek, risk analizi yapabilmek ve yeni pazar araştırmaları yürütebilmek için bu olaylardan gelen sinyalleri kullanırlar. Yatırımcılar ise piyasa olup biteni takip edebilmek, trendleri erken yakalayabilmek ve yatırımlarını doğru yere yönlendirip, etkin kararlar verebilmek için yapılan çıkarımlardan faydalanırlar. Ancak son yıllarda iyice artan finans haberleri bunları takip etmeyi ve bu metinleri işlemeyi imkansız hale getirmiştir. Bu sebeple finansal metinlerden otomatik çıkarım yapma ihtiyacı doğmuştur.

Finansal olay çıkarımı alanında İngilizce, Çince, Arapça gibi çeşitli dillerde uzun zamandır çalışmalar yapılmış ve halen de yapılmaya devam etmektedir, ancak bildiğimiz kadarıyla Türkçe için yapılmış bir çalışma bulunmamaktadır. Olay çıkarma görevi ilk zamanlarda genellikle örüntü eşleştirme yöntemleriyle gerçekleştirilmiştir. Kullanılan bu yöntem için alanındaki uzman kişiler belli kurallar ve şablonlar oluşturmuşlar ve örüntü eşleştirme ile metinlerde geçen olay ve argümanları bulabilmişler. Daha geniş kapsamlı çalışmalar gerçekleştirmek istediklerinde ise

örüntü eşleştirme ile yapılamayacağını anlayıp makine öğrenimi tekniklerini tercih etmişler. Ancak makine öğrenimi, görevi anlayabilmek ve doğru tahminleri yapabilmek için ona yol gösterecek işaretlere ihtiyaç duyar, bunlara etiketli veriler denilmektedir. Normalde etiketli veri elde edebilmek için uzmanlar tutulur ve işaretleme yapmaları istenir. Ancak insan eliyle yapılan bu işlemler oldukça zaman alan ve maliyetli bir iş olduğu için veri sayısı hep sınırlı kalmıştır. Var olan etiketli veriler kullanılarak makine öğrenimi eğitimleri gerçekleştirilmiş ancak hala veri sayısındaki eksiklikten ötürü istenen sonuçlara erişilememiştir. Bunun üzerine etiketli veriyi kullanmadan olay çıkarma denemeleri gerçekleştirilmiştir, ancak bunlar da etiketli verilerle eğitilen modellerin performansını yakalayamamıştır.

Eksik veri problemi sadece olay çıkarımı alanında değil daha bir çok makine öğrenimi çalışmalarında da ortaya çıkmıştır. Bu nedenle daha fazla nasıl etiketli veri elde edilebileceği üzerine çeşitli araştırmalar yapılmıştır. Bu çalışmalar kendi kendine eğitim, aktif öğrenme, öğrenme aktarımı ve zayıf denetim olarak dört grupta incelenebilirler. Kendi kendine eğitimde, etiketli veri kümesi üzerinde eğitilen bir model etiketsiz veri kümesindeki verilerin etiketlerini tahmin eder, ve bu yeni etiketi tahmin edilen veriler etiketli veri kümesine dahil edilir. Aktif öğrenme veri kümesindeki en büyük etkiyi yaratacak örneklerin bulunup bu örneklerin insan tarafından işaretlenmesiyle sağlanır. Bu sayede buraya koyulan efor önemli derecede düşer. Öğrenme aktarımı sıfırdan öğrenme yapmak yerine benzer bir görevde öğrenilenleri kullanmayı hedefler. Zayıf denetimde ise kesin olmayan, gürültülü kaynaklar kullanılarak etiketsiz verilere etiket atanması sağlanır. Bu kaynaklar kurallar, bilgi tabanları veya geliştirilen başka modellerin çıktıları olabilirler.

Bu tezle 3 ana katkıda bulunuyoruz;

1) Türkçe finans ve ekonomi alanında metin madenciliği uygulamaları geliştirmeye imkan sağlayacak bir referans veri kümesi sağlıyoruz. Bu derlem için 2010 – 2022 arasında farklı haber sitelerindeki şirketlere özel finans haberlerini toparladık. Olay çıkarımındaki alan kavramlarını ve aralarındaki ilişkileri tanımlayabilmek için ontoloji oluşturduk. Oluşturduğumuz ontoloji bu alanda yapılmış diğer akademik yayınların kavramları baz alınarak başladı, incelenen Türkçe haber metinlerinden yapılan çıkarımlarla listede eklemeler ve çıkarmalar yapılarak son haline getirildi. Toplandığımız şirket haberlerinden 600 tanesi seçilerek cümle ve kelimelerine ayrıldı. Daha sonra bu seçilen dokümanlar etiketlenmesi için işaretleme aracına verildi, ve ontolojiye uygun olarak olay tetikleyicileri ve olay argümanları işaretleyici tarafından etiketlendi.

2) Kendi sonuçlarımızı değerlendirmek ve ileride yapılacak olay çıkarımı çalışmalarına zemin hazırlamak için temel model sonuçlarımızı paylaşıyoruz. Bunu yapabilmek için elle etiketlediğimiz veri kümesini eğitim ve test olmak üzere ikiye ayırdık. Eğitim verilerini kullanarak eğittiğimiz farklı modellerin test verileri üzerindeki sonuçlarını karşılaştırdık ve en başarılı modeli temel olarak belirledik.

3) Denetimli makine öğrenimi modellerinin daha yüksek başarımlar elde edebilmesi ve görevi daha iyi öğrenebilmeleri için eğitim veri kümesini genişletecek çalışmalar yürüttük. Buradaki amacımız insan eforunu minimize ederek daha fazla etiketli veri elde edebilmektir. Bunun için aktif öğrenme ve zayıf denetim metotlarını denedik. Sadece aktif öğrenme metodunu kullanarak en az sayıda etiketli veri

ile temel model performansına erişebilmeyi hedefledik. Zayıf denetim tekniklerini kullanarak insan eforu koymadan etiketli veri elde etmeyi amaçladık. Bunun için iki fonksiyon geliştirdik, birincisi olayları belirten kelimelerin bulunduğu listeyi kullanarak cümlelerde olay olup olmadığını gösterir. Diğer fonksiyon ise eğittiğimiz model sonucunda çıkan tahmin ile var olan diğer gürültülü tahminleri entegre ederek son etiketli listeyi döner. Son olarak aktif öğrenme döngüsüne zayıf denetim fonksiyonlarını dahil ederek kaliteyi düşürmeden insan eforunu en aza indirmeye çalıştık.

Temel model için yaptığımız deneyler sonucunda eni iyi skoru BERT modeli ile elde ettik. Bu sonuçlar olay sınıflandırma için %85,89 ve argüman sınıflandırma için %65,81 makro ortalamalı F1 puanıdır. Sadece aktif öğrenme kullanıldığı durumda olay sınıflandırma için %30 efor kazanımı olurken yalnızca %0,06 performans düşüşü gözlenmiştir, argüman sınıflandırma için ise bu oranlar %25 ve %0,32 olarak kaydedilmiştir. Aktif öğrenme ve zayıf denetim metotlarının birlikte kullanıldığı durumda olay sınıflandırma için %54 efor kazanımı sağlanırken %0,7 oranında da performans iyileşmesi görülmüştür. Argüman sınıflandırma için bu oranlar %60 ve %2,32 olarak kaydedilmiştir. Son olarak elle etiketlenen verilerle zayıf denetim tarafından etiketleri atanan veriler birlikte kullanıldığında olay sınıflandırma için %2,91, argüman sınıflandırma için %13,76 lık bir performans artışı görülmüştür. Yaptığımız deneylerin sonuçlarına baktığımızda uyguladığımız metotların olumlu sonuçlar verdiğini, sadece daha fazla etiketli veri üretmekle kalmayıp performansı da arttırdığı görüyoruz.

Elle etiketlediğimiz veri kümesi ile diğer çalışmalarda oluşturulmuş veri kümelerini karşılaştırdığımızda derlemimiz cümle sayısı olarak diğerlerinin gerisinde olsa da geniş kapsamı sayesinde daha fazla olay içermektedir. Elde ettiğimiz son veri kümesindeki veri sayısının doküman, cümle, olay ve argüman başlıklarında elle etiketlediğimiz veri sayısını yaklaşık 50 ye katladığını görüyoruz.

Yaptığımız çalışmaların Türkçe finansal metinlerde yapılacak çalışmalara hız kazandıracağını ve daha fazla etiketli veriye ihtiyaç duyan görevlerde veri kümelerini genişletmede kullanılacağını düşünüyoruz.

TFEEC : TURKISH FINANCIAL EVENT EXTRACTION CORPUS

SUMMARY

With the developing world, the interest in digitization is increasing. Recently, digital transformation studies in various fields have also been works that emphasize the importance of this trend. Thanks to these, incredible numbers of texts are produced in digital environments. Although the extraction of information from these digital sources can be done manually for a small number of data, working with large amounts of data is very intensive and time consuming. In order to overcome these problems and automate manual work, information extraction techniques have been developed. Information extraction usually aims to automatically extract structured information from texts using natural language processing techniques. This makes the process more efficient by reducing human effort and costs in this area. Studies have gained popularity due to the benefits it provides.

One type of information extraction is event extraction. Event extraction is a challenging task that involves finding events in the text, identifying the types of events found, and identifying the corresponding arguments. Since it can produce structured information, it can contribute to the world of natural language processing by providing useful information for various tasks such as knowledge base building, question answering, etc. However, limited number of literature studies have been conducted in this area, and these studies generally focus on a specific area. One of these areas is studies in the domain of finance and economics.

Financial event extraction is very important for both companies and investors. Companies use signals from these events to get quick feedback on their products, conduct risk analysis and new market research. Investors, on the other hand, benefit from inferences in order to follow the general market trend, direct their investments to the right place and make effective decisions. However, the increasing financial news in recent years has made it impossible to follow and process these texts. For this reason, the need to make automatic extraction from financial texts has arisen.

In the domain of financial event extraction, studies in various languages such as English, Chinese and Arabic have been carried out for a while and still continue to be done, but to the best of our knowledge, there is no study conducted for Turkish. For the event extraction task, studies were carried out with pattern matching methods in the early times. Domain experts created certain rules and templates for the method used, and they were able to find the events and arguments in the texts by pattern matching. When they wanted to do more comprehensive studies, they preferred machine learning techniques. However, machine learning needs signals to guide it in order to understand the task and make accurate predictions, these are called labeled data. In order to obtain

the data on these labels, experts were hired and asked to annotate them. However, since these processes are very time-consuming and costly, the number of data has always been limited. Machine learning training were carried out using existing labeled data, but still the desired results could not be achieved due to the lack of data. Thereupon, event extraction experiments were carried out without using the labeled data. However, these are can not reach the performance of supervised machine learning models.

The missing data problem has emerged not only in the event extraction domain, but also in many machine learning studies. For this reason, various studies have been conducted on how to obtain more labeled data. These studies can be examined in four groups as self-training, active learning, transfer learning and weak supervision. In self-training, a model trained on the labeled dataset predicts the labels of the data in the unlabeled dataset, and this newly labeled predicted data is included in the labeled dataset. Active learning is achieved by finding the samples that will have the greatest impact in the dataset and annotating these examples by human. In this way, the effort put in here is significantly reduced. Transfer learning aims to use what has been learned in a similar task rather than learning from scratch. In weak supervision, on the other hand, it is provided to assign tags to unlabeled data by using imprecise, noisy sources. These resources can be rules, knowledge bases or outputs of other developed models.

We report three main contribution with this thesis.

1) We provide a gold standart dataset that will allow developing text mining applications in Turkish finance and economics domain. For this corpus, we have scraped company specific financial news from different news sites between 2010 and 2022. We created an ontology to define the field concepts in event extraction and the relationships between them. This ontology started on the basis of the concepts of other academic publications in this domain, and was finalized by making additions and deletions in the list with the inferences made from the analyzed Turkish news texts. 600 of the collected company-specific news were selected and separated into sentences and tokens. Then these selected documents were given to the annotation tool for labeling, then event triggers and event arguments were annotated manually according to the ontology.

2) We report the baseline model results so that we can evaluate our own results and form a baseline for future event extraction studies. To achieve this, we separate the manually labeled dataset into a training and test set. Using the training data, we compared the results of the different models we trained on the test data and determined the most successful model as a baseline.

3) We have carried out studies to expand the training dataset so that supervised machine learning models can achieve higher performance and learn the task better. Our aim here is to obtain more labeled data by minimizing human effort. For this, we tried active learning and weak supervision methods. By using only the active learning method, we aimed to reach the baseline model performance with the least number of labeled data. We aimed to obtain labeled data without human effort by using weak supervision techniques. We have developed two functions for this, the first one shows whether there are events in sentences using a list of words that indicate the events. The other function returns last labeled list by integrating the prediction resulting from the model we trained with other existing noisy predictions. Finally, we tried to minimize human

effort without reducing quality by including weak supervision functions in the active learning cycle.

As a result of our experiments for the baseline model, we obtained the best score with the Bert model. These results are a macro-averaged F1 score of 85.89% for event classification and 65.81% for argument classification. In the case where only active learning is used, 30% effort gain was observed for event classification, while only 0.06% performance decrease was observed. For argument classification these rates were recorded as 25% and 0.32%. When active learning and weak supervision methods are used together, 54% effort gain was achieved for event classification, and 0.7% performance improvement was observed. For the argument, these rates were recorded as 60% and 2.32%. Finally, when using manually labeled data and data assigned labels by weak supervision, a performance improvement of 2.91% for event classification and 13.76% for argument classification was observed. When we look at the results of our experiments, we see that the methods we applied gave positive results, not only producing more data labels, but also increasing performance.

When we compare the dataset that we labeled manually with the datasets created in other studies, although our corpus lags behind the others in terms of the number of sentences, it contains more events thanks to its wide scope. We see that the number of data in the last dataset we have obtained has increased to approximately 50 times the number of data we manually labeled in document, sentence, event and argument titles.

We think that our work will accelerate the studies to be done in Turkish financial text mining domain and will be used to expand datasets for the tasks that need more labeled data.

1. GİRİŞ

1.1 Tezin Amacı

Haberin günlük hayatımızda önemli bir yeri vardır. Gelişen teknoloji ve internet erişiminin kolaylığı ile artık daha fazla haber paylaşılmakta ve daha fazla okuyucu sayısına ulaşılmaktadır. İnsanlar gündemi, olayları, durum değişikliklerini takip etmek ve dünyadaki gelişmelerden kopmamak isterler. Tabii bunu sadece olayları takip etmek ve gelişmeleri öğrenmek için değil hayatlarında belli kararlar almak için de kullanırlar.

Özellikle finans ve ekonomi haberleri, şirketler ve bu alanda yatırım yapan/yapmak isteyen insanlar için daha önemlidir. Yatırım yapacak kişi ister uzman olsun isterse sıradan bir vatandaş fark etmez, önemli olan hızlı bir şekilde doğru habere ulaşmak ve haberden çıkarım yapabilmektir. Çünkü haberler bize piyasa eğilimleri ve çeşitli finansal araçların performansları hakkında bilgi verir. İnsanlar da burada edinilen bilgilerle doğru ve etkili bir şekilde yatırım yapma fırsatı yakalayabilirler. Ancak her gün dünyanın her yerinden yüzlerce belki de binlerce haber yazısı paylaşılıyor. Bu tonlarca haberi okumak ya da finansal anlamda en doğru bilgiyi yakalamak oldukça zor ve zaman isteyen bir süreçtir. Bu sebeple metinden önemli olayların/bilgilerin çıkarılması işi kritik bir hal almıştır.

Metinlerden otomatik olay çıkarımı yapmak çok kısa sürelerde çok iş yapmamızı sağlar ve bizi büyük bir yükten kurtararak hayatımızı kolaylaştırır. Doğal dil işleme dünyasında bu problemi çözmeye yönelik göreve olay çıkarımı denmektedir. Kısaca olay çıkarımı, doğal dil ile oluşturulmuş metinlerden olayı tetikleyen ifadeleri, olaya dahil olan argümanları ve bunların ilişkilerini otomatik olarak elde etmemizi sağlar.

Bu görev için gerçekleştirilen ilk çalışmalar genellikle örüntü eşleştirme yöntemi ile yapılmıştır. Bu yöntemde öncelikle alan uzmanları tarafından belirli kurallar ve şablonlar belirlenir, ardından örüntü eşleştirme ile olay ve argümanlar çıkarılır. Uzmanlar tarafından oluşturulan iyi tanımlanmış şablonlar sayesinde yöntemin

doğruluğu oldukça yüksek çıkmaktadır, ancak büyük ölçekli olay tasarımının maliyetli ve zaman alıcı olması nedeniyle sonraki çalışmalarda makine öğrenmesi tabanlı teknikler kullanılmıştır. Destek vektör makinesi (SVM), maksimum entropi (ME) gibi geleneksel makine öğrenmesi algoritmaları ve tekrarlayan sinir ağları (RNN), grafik sinir ağları (GNN), konvolüsyonel sinir ağları (CNN) gibi derin öğrenme yaklaşımları farklı alanlarda ayrı ayrı kullanılsa da yapılan çalışmaların temel yapıları birbirine benzemektedir. Her ikisinde de sınıflandırıcılar, etiketlenmiş veri kümesi üzerinde eğitilir ve metinden olayları çıkarmak için kullanılır. Ancak, yüksek performanslı bir sınıflandırıcıyı eğitmek, büyük bir veri kümesi gerektirir ve etiketli bir veri kümesi oluşturmak uzun, maliyetli ve oldukça yoğun emek gerektiren bir süreçtir. İngilizce, Çince, Arapça gibi farklı dillerde ekonomik ve finansal olaylar için etiketlenmiş veri kümeleri olmasına rağmen, maalesef Türkçe için daha önce böyle bir çalışmanın yapıldığından haberdar değiliz.

Biz de bu alandaki boşluğu doldurmak, Türkçe finansal ve ekonomik olay derlemi sunmak, ileriki araştırmalara temel hazırlamak için bu çalışmayı yaptık.

1.2 Tezin Katkıları

Bu tezle, üç ana katkı sağlıyoruz;

1. Bu çalışmadaki olayların kapsamını belirlemek için literatür taraması yaptık, ardından internet ortamında şirketlere özel haber metinlerini analiz ettik ve 25 tür ve 104 alt türden oluşan bir ontoloji oluşturduk. Bu ontolojiyi kullanarak, 600 haber makalesindeki olay tetikleyicilerini ve argümanlarını elle işaretledik. Bu oluşturduğumuz veri kümesini Türkçe ekonomi ve finans alanında denetimli makine öğrenimi modellerine etiketli veri sağlaması için herkese açık bir şekilde paylaştık.

2. Veri kümesini kullanarak olay çıkarma modelinin ön değerlendirmelerini yaptık. Gerçekleştirilen çalışmaların sonuçları, bu alanda yapılacak sonraki çalışmalar için bir temel sağlayacaktır.

3. Etiketlenmiş veri miktarını artırmak ve işaretleme sürecindeki insan eforunu en aza indirmek için aktif öğrenme (AL) ve zayıf denetim (WS) yöntemlerini denedik, yapılan çalışmaların ilk sonuçlarını değerlendirdik ve otomatik olarak elde ettiğimiz yeni

etiketli verileri referans kümemizdeki verilerle birleştirerek derlemimizi genişlettik. Bu genişletilmiş yeni derlem ile yaptığımız deney sonucunda, elle etiketlenen ve otomatik olarak etiketi atanan veri kümelerini birlikte kullandığımız durumda, modelin çok daha iyi performans sonucu verdiğini gördük. Bu da insan eforunu minimize ederek yüksek kaliteli yeni etiketler elde edebildiğimizi gösteriyor.

1.3 Genel Bakış

Tezimizi oluşturan bölümler aşağıdaki gibi sıralanmıştır.

1. Bölümde genel bilgilerden bahsedilmekte, çalışmanın amaçları, kazanımları hakkında bilgi verilmektedir.
2. Bölümde şu ana kadar oluşturulmuş olay çıkarım derlemleri ve daha fazla etiketli veri elde etmek için yapılan çalışmalar incelenmektedir.
3. Bölümde olayın tanımı, doğal dil işleme görevi olarak olay çıkarımı ve finansal olay çıkarımı örnekleri gösterilmektedir.
4. Bölümde veri kümesini oluşturma süreci hakkında bilgi verilmektedir.
5. Bölümde kullandığımız metotlar anlatılmaktadır.
6. Bölümde çalışmamıza yaptığımız testler ve değerlendirmeleri yapılmaktadır.
7. Bölümde sonuçlarımız ve çalışmanın faydaları vurgulanmaktadır.

2. LİTERATÜR ARAŞTIRMASI

2.1 Olay Çıkarımı Derlemi

Veri etiketleme, makine öğrenimi modelinin tahmin etmesini istediğimiz sonucu göstererek doğru tahminde bulunmasını sağlama sürecidir. Verileri işaretleme süreci tarafından oluşturulan eğitim verileri, verilen görevleri doğru bir şekilde anlamak ve istenen tahminleri yapmak için modellere rehberlik eder. Bu nedenle, model başarısı için bir derlem oluşturmak esastır. Ancak bu göründüğü kadar kolay değildir, bu bölümde yapılmış olan ilgili çalışmaları gözden geçireceğiz.

Normalde oluşturulması istenen veri kümesi alan bilgisine sahip uzmanların elle işaretlemesiyle yapılır. Ancak bu onlar için zaman alıcı ve oldukça yoğun emek gerektiren bir iştir, bu nedenle derlem oluşturma üzerine yapılan ilk çalışmalar her zaman daha küçük ve dar kapsamlı olmuştur. Örneğin ACE 2005 derlemi [1] İngilizce, Arapça ve Çince dilleri de dahil olmak üzere toplam 599 belge içerir ve toplamda 8 tür ve 33 alt tür bulundurur. Benzer bir çalışma olan TAC-KPB 2015 derlemi [8], yalnızca İngilizce olarak 360 belgeye sahiptir ve 9 tür ve 38 alt tür içerir. Son yıllarda yapılan çalışmalarda bu alandaki veri kümelerinin kapsamı genişletilmiş ve daha fazla etiketli veriler içeren derlemler sunulmuştur. SENTIVENT derlemi [12] İngilizce olarak 288 belge içermektedir ve bu çalışma için 18 olay türü ve 64 alt türü tanımlanmıştır. Benzer bir çalışma olarak Türkçe için FrameNet [42] oluşturulmuştur. Ancak FrameNet yüklemle ilgili ilişkileri inceler ve finan ve ekonomi alandaki ifadeleri içermez.

Çalışmamız için bu alanda yapılan diğer çalışmalardaki olay tetikleyici ve olay argümanı etiketleme/işaretleme süreçlerini benimsedik. Bununla birlikte diğer çalışmalardaki olay kapsamlarını inceledik, ancak bu çalışmalarda kullanılan türler ve alt türler bizim için yeterli olmadığından okuduğumuz haberlerde belirlediğimiz yeni olay türleri ve alt türleri ile bu kapsamı genişleterek yeni bir kapsam oluşturduk.

Yapılan çalışmalar kapsam ve büyüklük açısından bizi iyi bir seviyeye getirse de veri sayısı anlamında belirli bir sınırı aşmamıza izin vermemektedir. Eksik veri sorunundan etkilenmemek için denetimsiz öğrenme modelleri [39]–[41] de önerilmiştir. Ancak yapılan çalışmaları karşılaştırdığımızda denetimli öğrenme modellerinin denetimsiz öğrenme modellerinden çok daha iyi performans gösterdiğini görmekteyiz. Bu da etiketli veriye olan ihtiyacı gözler önüne serer.

2.2 Veri Genişletme Metotları

Makine öğrenimi algoritmaları, olayları, karmaşık kalıpları öğrenmek ve daha iyi genelleme yapmak için büyük miktarda veriye ihtiyaç duyar. Çünkü daha fazla veri, modeli daha iyi anlamaya yardımcı olur ve başarılı tahminler yapma olasılığını artırır. Normalde geleneksel yaklaşım, verileri etiketlemek için uzmanları işe almaktır, ancak etiket eklenecek veri kümesi daha büyükse, süreç çok pahalı ve zor hale gelir. Veri seyrekliğini önlemek için, küçük etiketli veri kümesi veya sözlükler kullanan ve daha fazla etiketli veri elde eden modeller geliştirebiliriz. Bu bölümde, daha fazla etiketli veri elde etmek için geliştirilmiş yöntemlere göz atacağız.

2.2.1 Kendi kendine eğitim

Bu strateji için etiketli ve etiketsiz veri kümeleri mevcuttur. Etiketli veri kümesindeki verilerle bir model eğitilir ve ardından etiketlenmemiş örnekleri sınıflandırmak için kullanılır. Bu tahmin edilen etiketli veriler eğitim verilerine dahil edilir ve model son eğitim verileriyle yeniden eğitilir, süreç etiketsiz veri kalmayana kadar tekrarlanır [16,17]. Bu yöntem incelendiğinde tek başına kullanılmasının performans açısından yeterli olmadığı görülmüştür.

2.2.2 Aktif öğrenme

İyi bir model oluşturmak için kullanılan veriler aynı etkiyi göstermez. Aktif öğrenme yaklaşımının amacı, mevcut tüm verileri kullanmaktan ziyade bilgi açısından en değerli örnekleri seçmektir. Seçilen en değerli örnekler işaretlemesi için etiketleme yapacak kişilere verilir, yeni etiketleri atanan veriler etiketli veri kümesine dahil edilirler. Aktif öğrenme, varlık ismi tanıma [18], metin sınıflandırma [19], sözcük

türü etiketleme [20], sözcük anlam belirsizliği giderme [22] ve hatta olay çıkarma [23] gibi birçok görevde başarıyla uygulanmıştır. Biz de işaretleyeceğimiz verileri önceliklendirmek ve kişilerin ortaya koyduğu eforu düşürebilmek için bu çalışmada aktif öğrenmeyi kullanmayı planladık. Diğer çalışmalardan farklı olarak aktif öğrenme döngüsüne zayıf denetim fonksiyonlarını ekleyerek denemelerimizi de gerçekleştirdik.

2.2.3 Öğrenme aktarımı

Normalde makine öğrenmesi için sıfırdan öğrenme yapılır. Ancak bir görevde veya konuda öğrenilenleri başka bir görevde kullanmak avantajlı olacağından, diğer görevlerden elde edilen bilgiler ana hedefteki görevin çözümü için kullanılır. Öğrenme aktarımı ile önceki bilgiler kullanılarak daha az eğitim verisi ile daha yüksek başarı gösteren ve daha hızlı öğrenen modeller elde edilir. Bu avantajından dolayı birçok olay çıkarımı [24]–[26] çalışmasında bu yöntem tercih edilmiştir. Ancak Türkçe ekonomi ve finans alanında olay çıkarma konusunda etiketli veri veya eğitilmiş bir model olmadığı için bu yöntemden yararlanamıyoruz.

2.2.4 Zayıf denetim

Bu yaklaşım, kesin olmayan bilgi kaynaklarını kullanarak etiketlenmemiş verileri etiketlemek için göstergeler sağlar. Zayıf denetim kullanarak etiketli veriler elde etmek için araştırmacılar, bilgi tabanları [13,31,32], sezgisel kurallar [27] ve alanında uzmanlaşmış kişilerin yardımını alarak fonksiyonlar oluştururlar. Bu, büyük miktarlarda daha ucuz, etiketlenmemiş verilerden yararlanmayı mümkün kılar ve çoğu çalışma, genellikle gürültülü ve kalitesiz olmalarına rağmen, model eğitimi için zayıf etiketlerden faydalanır. Bu alanda çeşitli [33] sistemler de geliştirilmiş olup, küçük kod parçacıkları yazarak etiketlenmemiş verilerin etiket kümelerini elde etmemizi sağlarlar.

3. OLAY ÇIKARIMI

Bu bölümde genel olarak olayın finansal alanda nasıl var olduğunu açıklayacak ve bu olayları gösteren ifadeleri, olaya katılan argümanları tarif edeceğiz.

3.1 Olay

Olay "Olmak" fiilinden türetilmiş bir kelimedir, bu kelimeye yakın veya eş anlamlı olan diğer sözcüklere örnek olarak "Hadise", "Vaka", "Aksiyon" verilebilir.

Olay, birden fazla katılımcıyı içeren belirli bir durumun oluşmasının, ortaya çıkmasının, gerçekleşmesinin veya bu durumda bir değişiklik yaşanmasının metinsel bir açıklamasıdır. Genel olarak olay, tanıklık edilen, insanların ilgisini cezbeden işler olarak tanımlanabilir. Bu durumlar geçmişte, gelecekte veya şimdiki zamanda olabilirler. Doğası gereği geçicidirler, uzayda belirli bir zamanda belirli bir süreyle yer işgal ederler ve ne olduğunu, kimin dahil olduğunu, hangi zamanda ve nerede olduğunu açıklarlar.

Finansal alandaki olay ise çalışmamızın kapsamını oluşturmaktadır. Bu doğrultuda finansal olay, haber metinlerindeki şirketleri etkileyen veya dahil eden hadiselerdir. Şirkete özel haberlerin birincil amacı, işletmeler ve ekonomi ile ilgili mevcut durumdaki değişiklikleri bildirmektir. Anlaşmalar, çalışan değişiklikleri, ürün lansmanları, şirket birleşmeleri, davalar, vb. ekonomi haberlerinde nelerin bir olayı oluşturduğuna dair örneklerden bazılarıdır.

Olay çıkarma, çok sayıdaki belgelerden olayları, varlıkları hızlı bir şekilde tanımlamayı hedefleyen doğal dil işleme görevidir. Metindeki belirli olay türlerinin örneklerini ve bunlarla ilişkili argümanları tanımlamayı ifade eder. Her olay, o olayı çağrıştıran bir kelime veya bir metin aralığı, olay tetikleyicisi (çoğunlukla tek fiiller veya deyimsel fiiller, aynı zamanda isimler, deyimsel isimler, zamirler ve zarflar) ile temsil edilir. Metinde ilk olarak olayın varlığı sözlüksel bir tetikleyici ile belirtilir.

Sonra, her olaya belirli bir tür atanır. Tetikleyicilerin tespiti ve sınıflandırılmasından sonra olayın argümanları bulunur. Olay argümanları, olayda katılımcı olarak yer alan varlıklar veya değerlerdir. Olaya katılan argümanları bulmak için olaya atıfta bulunan ifadeler çıkartılır (kişiler, şirketler, ürünler, vb.) ve bunlara uygun argümanlar atanır.

Bu iki ana alt görev, olay tetikleyicisi bulma ve argüman tespit etme, birbirlerine oldukça bağlıdırlar. Tetikleyici tanımlama genellikle bağımsız bir görev olarak ele alınır ve argüman bulma ve özellik atamalar tetikleyici tanımlamanın sonuçlarından etkilenir.

Örneğin, metindeki bir olayı aşağıdaki bileşenlerle ifade edebiliriz.

- *"Türkiye'nin inşaat firmalarından Ataç İnşaat ve Sanayi A. , Kuzey Irak'ta baraj inşa etmek üzere geçtiğimiz hafta **anlaşma** imzaladı ."*

Olay tetikleyici : Metindeki olayı en açık şekilde ifade eden kelimedir. Bu cümledeki olay türünü belirleyen *"anlaşma"* kelimesidir.

Olay argümanı : Olay metninde bir katılımcı veya özellik olarak yer alan ve olay tetikçisine atıfta bulunan varlık, zamansal ifade veya değerdir. Cümledeki *"Ataç İnşaat ve Sanayi A."* şirket olarak, *"geçtiğimiz hafta"* ise zaman ifadesi olarak yer almaktadır.

İlerleyen bölümlerde, bu konuları daha ayrıntılı olarak açıklayacağız.

3.2 Olay Tetikleyicisi

Olay algılama, olay çıkarmanın önemli ve oldukça zorlu bir alt görevi olarak kabul edilir ve metindeki belirli olay türlerinin tanımlandırılmasını ve bunları tam olarak olay türlerine göre sınıflandırılmasını içerir. Aşağıdaki örneklerde olay tetikleyicilerini fark edilebilmesi açısından kalın harflerle gösterdik. Ayrıca hangi ana tür ve alt türe ait olduklarını da altlarında belirttik.

- *"İki Suudi bankasından National Commercial Bank ile Riyad Bank arasındaki 200 milyar dolarlık **birleşme** iptal edildi ."*
  - Olay Türü, Alt Türü [*birleşme*] : Birleşme

- *"Credit Suisse Erdemir için **fiyat hedefini** 2.10 liradan 2.30 liraya yükseltti ."*
  - Olay Türü, Alt Türü [*fiyat hedefinin*] : Hisse_Pay, Fiyat_Hedefi
- *"Credit Suisse 4. çeyrekte **zarar** etti ."*
  - Olay Türü, Alt Türü [*zarar*] : Zarar, Duyuru
- *"Credit Suisse , kıta Avrupası ve ABD için **tavsiyesini** ise Ağırlık Azalt olarak sürdürürken , gelişmekte olan piyasalar için Ağırlık Artır tavsiyesini korudu ."*
  - Olay Türü, Alt Türü [*tavsiyesini*] : Analist, Tavsiye

Bu olayları tetikleyen ifadeler tek bir sözcük tipine ait olmak zorunda değiller, bu ifadeler fiil, isim, zamir veya zarf da olabilirler.

- *"Denizbank 1.400 kişiyi **işe alacak** ."*
  - Olay Türü, Alt Türü [*işe alacak*] : Çalışan, İstihdam
  - Sözcük Tipi : Fiil
- *"Marmara Bölgesi'ndeki 59 mağazasıyla faaliyet gösteren KİM Marketler Zinciri, 2012 yılında 35 yeni mağaza açarak 400 kişiye **istihdam** sağlayacak. "*
  - Olay Türü, Alt Türü [*istihdam*] : Çalışan, İstihdam
  - Sözcük Tipi : İsim
- *"Paşabahçe'nin aktiflerinde bulunan bazı gayrimenkullerin kısmi **bölünme** suretiyle devralınmasına izin çıktı . "*
  - Olay Türü, Alt Türü [*bölünme*] : Kurum, Bölünme
  - Sözcük Tipi : Zarf

Olayın algılanma aşaması her zaman net değildir. Birden fazla sözcüğün tetikleyici olarak işaretlenebileceği durumlar da vardır. Tetikleyici ifadeler ilgilenilen olayı veya yazarın ilgilenmediği başka bir olayı, durumu temsil edebilir.

- *"LONDRA - Morgan Stanley tarafından bugün yayınlanan bir raporda bazı Türk hisseleri için yatırım **tavsiyelerini** ve hedef fiyatlarını değiştirdi ."*
  - Olay Türü, Alt Türü [*tavsiyelerini*] : Analist, Tavsiye
  - Yatırım kelimesi bu cümlede bir yere yatırım yapmak anlamında kullanılmamıştır, sadece tavsiyeyi betimler.

- *"Finans Yatırım bir hisse için fiyat hedefini ve yatırım **tavsiyesini** değiştirdi ."*
  - Olay Türü, Alt Türü [*tavsiyesini*] : Analist, Tavsiye
  - Cümlenin başındaki yatırım kelimesi bir eylem değil, şirketin ismidir[*Finans Yatırım*].

Olayların tetikleyicisinin belirlenmesi, şimdiye kadar gördüğümüz örneklerde olduğu gibi genellikle basittir, ancak her zaman bu kadar basit olmamaktadır. Tetikleyici olarak işaretlenebilecek birden fazla sözcük bulduğumuz durumlar olabiliyor. Veya tetikleyici olarak tek bir sözcük seçmenin doğru/yeterli olmadığını düşündüğümüz olay türleri ile de karşılaşabiliyoruz. Genel prensibimiz gereği, olay türünü veya alt türünü anlamını kaybetmeden olabildiğince küçük tutmaya çalışıyoruz. Ancak bu örneklerde yer alan ifadeler yalnızca bir araya geldiklerinde bir olayı belirtmiş oluyorlar. Bu ifadeler yan yana oldukları gibi birbirinden uzakta da olabilir ve hatta cümle içindeki yerleri de değişebilmektedir.

- *"Bu segmentteki **gelirler** , yılın ilk yarısında 2011'ün aynı dönemine oranla yüzde 66 **düştü** ."*
  - Olay Türü, Alt Türü [*gelirler...düştü*] : Gelir, Azalış
- *"% 48,04 oranında hissesine sahip olduğumuz iştirakimiz GENPAZ ORMAN ÜRÜNLERİ TİCARET VE PAZARLAMA A.Ş. 16.650.000,00 TL olan tamamı ödenmiş **sermayesini** 10.000.000,00 TL **azaltarak** 6.650.000,00 TL ye indirmiş ve diğer yasal işlemlerini tamamlamıştır ."*
  - Olay Türü, Alt Türü [*sermayesini...azaltarak*] : Sermaye, Azaltma

- *"Almanya'nın bayrak taşıyıcı hava yolu şirketi Lufthansa , 2016 yılında net **karını** yüzde 4,6 **yükseltti** ."*
  - Olay Türü, Alt Türü [*karını...yükseltti*] : Kar, Artış

Bir cümlede bir sadece bir olay olmak zorunda değildir. Aynı cümlede birden fazla olay da anlatılabilir. Bunlar genellikle karmaşık veya uzun cümlelerde yada bir yan tümceden oluşan basit cümlelerde karşımıza çıkarlar.

- *"Haziran ayında gerçekleşen satış sonrası **devir** işlemlerinin tamamlanmasıyla bankanın da **ismi** ve **logosu değiştirildi** ."*
  - Olay Türü, Alt Türü [*devir*] : Hisse_Pay, Devri
  - Olay Türü, Alt Türü [*ismi...değiştirildi*] : Bilgi_Değiştirme, İsim_Değiştirme
  - Olay Türü, Alt Türü [*logosu değiştirildi*] : Bilgi_Değiştirme, Logo_Değiştirme
- *"Son bir yıl içinde sürekli olarak hisse **devri** sözleşmeleri açıklayan Ericom son olarak hem yeni şirket satın **alıyor** , hem de yine **satış** açıklıyor ."*
  - Olay Türü, Alt Türü [*devri*] : Hisse_Pay, Devri
  - Olay Türü, Alt Türü [*alıyor*] : Hisse_Pay, Alış
  - Olay Türü, Alt Türü [*satış*] : Hisse_Pay, Satış
- *"İngiltere'nin 3 büyük bankası HSBC , RBS ve Barclays döviz kuru manipülasyonu **skandalında** 924 milyon dolar daha **ceza** ödeyecekler ."*
  - Olay Türü, Alt Türü [*skandalında*] : Marka, Skandal
  - Olay Türü, Alt Türü [*ceza*] : Yasa, Ceza

3.3 Olay Argümanı

Olaylar, değişen durumu tanımlar ve eylemleri, ilişkileri, oluşumları vb. kapsar. Bunlar olaya şu veya bu şekilde dahil olan birden çok katılımcı varlığı içerir. Her olaya karşılık gelen bir dizi argümanı vardır.

Bu bölüm, argüman bulma, yani belirli bir olay türüne sahip bir tetikleyiciye karşılık gelen argümanları tanımlamaya odaklanır. Bu görev, tetikleyiciyi bulma ve onu sınıflandırma görevinden sonra ikinci bir adım olarak gelir. Teoride, iki görev birbirine son derece bağımlıdır, ancak pratikte, iki ayrı adım olarak uygulanırlar: ilk önce tetikleyici tanımlama bağımsız bir görev olarak ele alınır ve argüman bulma ve nitelik atamalarının her biri tetikleyici tanımlamanın sonuçlarına bağlıdır.

- *"Chevron , geçen yılın aynı döneminde 2,04 milyar dolar net* ***kara ulaşırken*** *, bu senenin üçüncü çeyreğinde şirketin net* ***karı*** *yüzde 37,2 düşüşle 1,28 milyar dolara* ***geriledi*** *."*

Olay tetikleyici : Cümlede olayı anlatan *kar* ifadeleridir, bunlar olay türünü belirlerler, devamında bulunan *ulaşan* ve *geriledi* ifadeleriyse olayın alt türlerini belirtir. *kara ulaşmak* ifadesinin olay karşılığı tür olarak kar, alt tür olarak artıştır. Devamında gelen tetikleyici ifade *karı geriledi* için türümüz yine kar, alt türümüz ise azalıştır.

Olay argümanı : *Chevron* olayların sahibi olan argümandır ve rolü şirkettir. *geçen yılın aynı döneminde* ifadesi karın gerçekleştiği zamanı belirtir, rolü zamandır. *2,04 milyar dolar* net kar miktarıdır, rolü paradır. Devamında gelen *bu senenin üçüncü çeyreğinde* ifadesi ise diğer olayın bir argümanıdır ve yine karın gerekleştiği dönemi ifade eder, rolü zamandır. *yüzde 37,2* kardaki düşüş oranını vermektedir, rolü orandır. Ve son olarak *1,28 milyar dolara* ifadesi son durumdaki kar miktarını belirtmektedir, rolü de paradır.

- *"Biteks , polyester ve viskondan iplik **üretiyor** ."*
  - Olay Türü, Alt Türü [*üretiyor*] : Ürün, İmalat
  - Argüman [*Biteks*] : Şirket
  - Argüman [*iplik*] : Ürün
- *"Şirketin CEO'luğunu yürüten Cenk Alper , Sabancı Holding Sanayi Grup Başkanlığı'na **atandı** . "*
  - Olay Türü, Alt Türü [*atandı*] : Çalışan, Atama
  - Argüman [*CEO'luğunu*] : Ünvan
  - Argüman [*Cenk Alper*] : Kişi
  - Argüman [*Sabancı Holding Sanayi Grup Başkanlığı'na*] : Ünvan
- *"Sabancı Holding Aksigorta **sermayesindeki payı** 23 Mart itibariyle yüzde 35.33 sınırına ulaştı ."*
  - Olay Türü, Alt Türü [*sermayesindeki payı*] : Sermaye, Payı
  - Argüman [*Sabancı Holding Aksigorta*] : Şirket
  - Argüman [*23 Mart itibariyle*] : Zaman
  - Argüman [*yüzde 35.33*] : Oran
- *"Buna göre , şirketin geçen senenin dördüncü çeyreğinde elde ettiği **gelir** bir önceki yılın aynı dönemine kıyasla yüzde 16 **azalışla** 21,9 milyar dolara **düştü**."*
  - Olay Türü, Alt Türü [*gelir...azalış...düşme*] : Gelir, Azalış
  - Argüman [*geçen senenin dördüncü çeyreğinde*] : Zaman
  - Argüman [*bir önceki yılın aynı dönemine*] : Zaman
  - Argüman [*yüzde 16*] : Oran
  - Argüman [*21,9 milyar dolara*] : Para

- *"Eczacıbaşı Yapı Grubu'nun Rusya'da **kurduğu** 37 milyon avroya mal olan karo ve seramik fabrikasının açılışı dün düzenlenen törenle yapıldı ."*
  - Olay Türü, Alt Türü [*kurduğu*] : İş_Alanı, Açma
  - Argüman [*Eczacıbaşı Yapı Grubu'nun*] : Şirket
  - Argüman [*Rusya'da*] : Yer
  - Argüman [*37 milyon avroya*] : Para
  - Argüman [*fabrika*] : İş Alanı

4. VERİ KÜMESİ OLUŞTURMA

4.1 İşaretleme Süreci

Türkçe finans ve ekonomi haberlerinden olayların çıkarılması görevine eğitim kümesi sağlamak amacıyla bir referans veri kümesi oluşturmak istedik. Bunun için önce olayları ve ilişkilerini tanımlayabilmek adına ontoloji oluşturduk.

Ontoloji, alandaki kavramları, aralarındaki ilişkileri ve özelliklerini tanımlamamızı sağlar. Ekonomik olayların tespitinde rehberlik etmesi için bir ontoloji tasarladık, buradaki amacımız, bu alandaki bilgi paylaşımını sağlamak, ortak anlayışı oluşturmak, yeniden kullanımını ve gelişimini desteklemektir. Çalışmanın başında, finansal ve ekonomik alanda olay çıkarımı üzerine çalışan diğer makaleleri gözden geçirdik. Ardından şirketlere ait çok sayıda haber makalesi okuduk. Diğer çalışmalardan elde ettiğimiz olay kümesini okuduğumuz haberlerdeki olayların sıklığına, uyumuna ve farklılığına göre düzenledik, eklemeler ve çıkarmalar yaptık. Sonuç olarak 25 olay tipi ve 104 alt tipi olan bir ontoloji elde ettik. Olay türlerinin ve alt türlerinin listesi Çizelge 4.1'de görülebilir.

Veri kümesine eklediğimiz dokümanları www.borsagundem.com sitesinde 2010-2022 yılları arasında yayınlanan şirket haberlerinden aldık. Bu dokümanlardan oluşan derlemimiz 34.746 haber makalesi, 323.945 cümle ve 6.180.243 kelime içermektedir.

Genelde haberler toplanırken belli listelerdeki şirketler filtrelenir, diğer şirket haberleri göz ardı edilir. Bunun bir amacı çeşitlilik sağlamak bir diğeri de bu şirketler hakkında daha fazla haber buluyor olmak, yani frekans açısından daha anlamlı olması. Benzer şekilde biz de haberleri tararken öncelikle Capital 500'deki şirketlerin haberlerini seçmek istedik, ancak oluşturduğumuz ontoloji geniş kapsamlı olduğu için bu şirketlere ait bazı olay türlerine ilişkin örnekler bulamadık. Örneğin şirket kuruluşu, şirketin kapanması veya iflası gibi haberler büyük şirketler için çok rastlanan olaylar

Çizelge 4.1 : Olay tür ve alt tür listesi.

Olay Türü	Olay Alt Türü
Analist	Beklenti, Raporlama, Tavsiye
Anlaşma	
Bilgi Değiştirme	İsim Değiştirme, Logo Değiştirme
Birleşme	
Borç	
Borsa	Halka Arz, İşlem Görme, Pazara Girme, Pazardan Çıkma
Çalışan	Atama, Değişiklik, Grev, İş Sözleşmesi, İşten Çıkarma, İstifa, İstihdam, Kaza, Ölüm
Finansal Sonuç	Dönem Bilançosu, Kredi Notu, Mali Tablo, Satış Hacmi
Gelir	Artış, Azalış, Duyuru
Gider	Artış, Azalış, Duyuru
Hisse Pay	Alış, Devri, Dönüşüm, Fiyat Hedefi, Geri Alma, Rehin, Satış
İhale	
İhracat	
İş Alanı	Açma, Alış, Kapama, Satış
Kar	Artış, Azalış, Duyuru
Kurum	Açma, Bölünme, İflas, Kapama, Yeniden Yapılanma
Macroekonomi	Devalüasyon, Endeks, Enflasyon, Faiz, İşsizlik, Milli Gelir, Pazar Payı, Zam
Marka	Etik, Hibe, İmaj, Kadına Şiddet, Ödül, Reklam, Rüşvet, Skandal, Sosyal Sorumluluk, Spekülasyon, Sponsor, Üyelik, Yolsuzluk
Proje	
Sermaye	Arttırma, Azaltma, Payı, Tavanı
Temettü	
Ürün	Alma, Çeşitlendirme, İmalat, Lansman, Satma, Üretime Ara Verme, Üretime Başlama, Üretime Devam Etme, Üretime Son Verme
Yasa	Beraat, Ceza, Dava, İhlal, İtiraz, Soruşturma, Suç Duyurusu, Tazminat, Tedbir, Uzlaşma, Yargılama, Yasal İşlem
Yatırım	
Zarar	Artış, Azalış, Duyuru

olmadığı için bu türlerde örneğe rastlamadık. Çalışmamızın geniş kapsamlı olmasını istediğimiz için şirket seçimi konusunda herhangi bir kısıtlama yapmama kararı aldık.

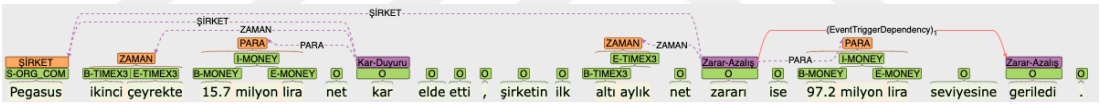
İşaretleme sürecine geçmeden önce ön işlem olarak sonraki adımlarda bize fayda sağlayacak bir [34,36] model çalıştırdık. Cümlelerdeki varlık adı tanıma ve zamansal ifadeleri bulabilen bu model girdi olarak aldığı dokümanları cümle ve kelimelere ayırır, bulduğu etiketlerle birlikte çıktısını verir.

Bu model sayesinde bulabildiğimiz varlıklardan bazıları aşağıdaki gibidir:

- Kişi
- Sayı
- Tarih
- Konum
- Oran
- Ünvan
- Organizasyon
- Para
- vb.

Bu modelden aldığımız çıktıdan rastgele 600 dokümanı seçiyor ve işaretleme aracına girdi olarak veriyoruz. Bu çalışmada yapılan tüm etiketleme işlemleri WebAnno işaretleme aracı [35] kullanılarak gerçekleştirilmiştir ve işaretlemeler cümle düzeyinde gerçekleştirilmiştir.

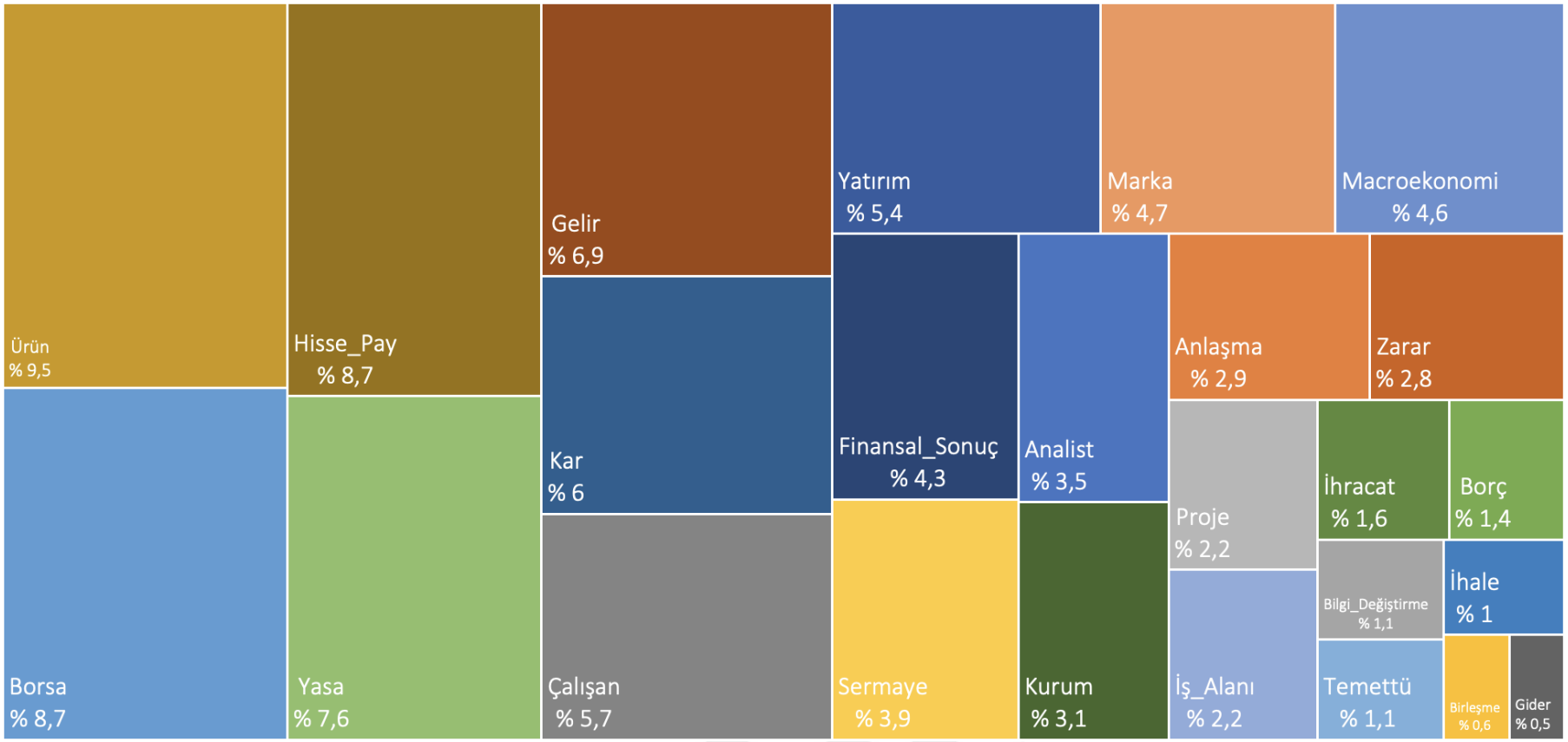
Bu dokümanlarda olayları gösteren tetikleyici ifadeler, olayı tarifleyen veya olaya katılan argümanlar ve bunların ilişkileri işaretlenmiştir. Örnek işlem Şekil 4.1'de görünmektedir. Yeşil etiketler ön işlem sonucunda elde ettiğimiz etiketleri göstermektedir. Turuncu (olay argümanı) ve mor(olay tetikleyicisi) etiketleri ise işaretleme sürecinde seçilmiştir.



Şekil 4.1 : İşaretlenmiş örnek bir cümle.

Veri kümemizdeki belgelerden çıkarılan olay türlerinin sıklık dağılımı Şekil 4.2'de gösterildiği gibidir.

Ontolojide yer alan olay türleri, alt türleri ve olay argümanlarının listesi https://github.com/kadirsinas/TFEEC adresinde mevcuttur. Buna ek olarak, belgelerden çıkarılan olay türlerinin sıklık dağılımı, veri seti ile ilgili istatistiksel bilgiler, işaretlemelerin nasıl yapıldığına ilişkin talimatlar ve işaretlemelere ait örnek ekran görüntüleri de bulunmaktadır.



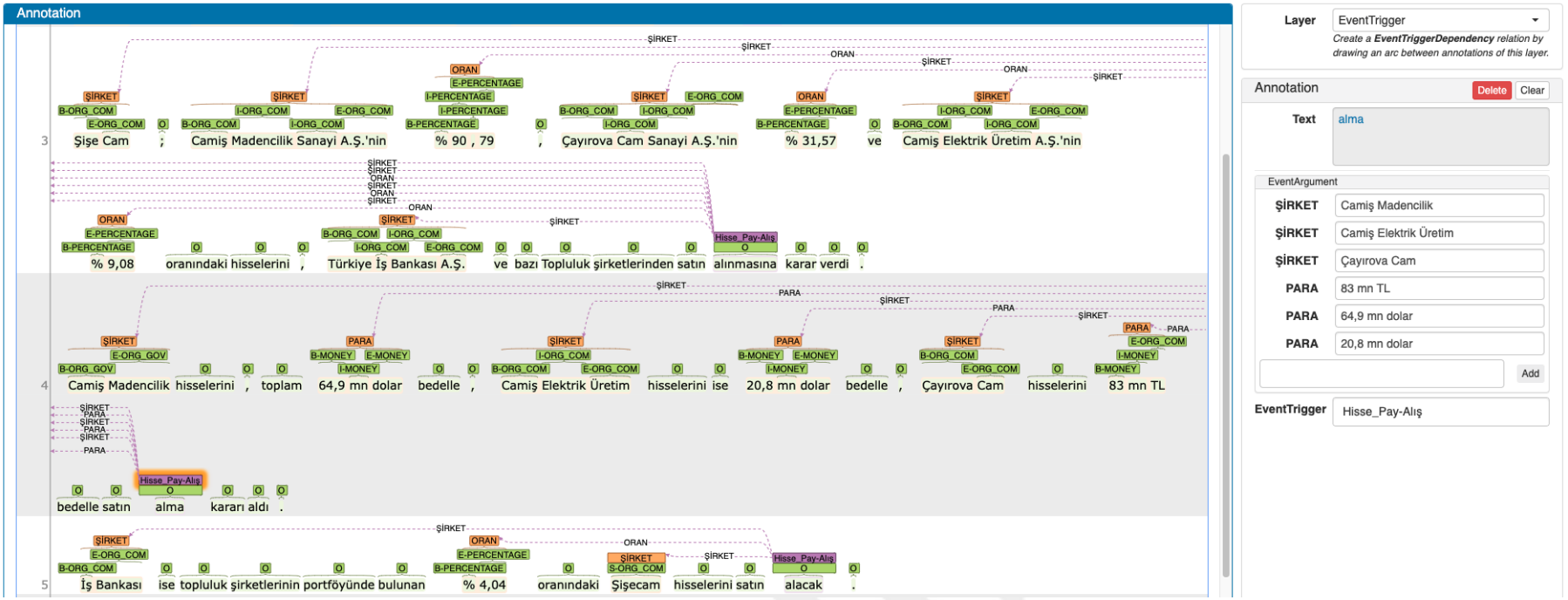
Şekil 4.2 : Türkçe finansal olay veri kümesindeki olay dağılımı.

4.2 İşaretleme Aracı

Metinleri işlemek, onları etiketlemek ve çıktılarını belli bir formatta oluşturmak oldukça meşakkatli bir iştir. Ancak bu işlemleri bizim için kolaylaştıran araçlar geliştirilmiştir, bunlara etiketleme araçları denmektedir. Bu araçlar, metinlere veya görüntülere etiket atamamıza, bilgi eklememize yardımcı olur, işaretleme sürecini ve kullanıcı deneyimini iyileştirirler.

Biz de çalışmamızdaki tüm işaretleme işlemleri için online metin etiketleme platformu olan WebAnno'yu tercih ettik. Bu platform üzerinde olay tetikleyici, olay argümanı ve varlık isimlerinin etiketlerini sisteme girip etiketleme işini kullanıcı dostu bir hale getirdik. İşaretleme sonunda ise formatlı bir şekilde çıktı alabildik.

Şekil 4.3'te platformun arayüzü ve örnek işaretlemeler görünmektedir.



Şekil 4.3 : WebAnno işaretleme platformu.

5. VERİ KÜMESİ GENİŞLETME YÖNTEMİ

5.1 Aktif Öğrenme

Günümüzde daha fazla veriye ulaşmak kolaylaştı, artan paylaşım sayıları sayesinde daha fazla içerik internet ortamında ulaşılabilir oluyor. Bu verilerin toplanması ve depolanması da artık daha kolay hale gelmiştir, ancak bunlar etiketlenmemiş olduğundan makine öğrenimi için sorun teşkil etmektedir. Bu verilerin elle etiketlenmesi gerekir fakat veri miktarı elle etiketlemek için çok büyüktür ve ilgililerin bu verilerle iyi modeller eğitmesi oldukça zor hale gelir. Bu tür durumlarda verileri akıllı bir şekilde işaretlemek için bazı önceliklendirmelerin yapılması süreci verimli hale getirir. Bu sayede süreç modelin mevcut tüm eğitim verileri üzerinde eğitilmesi yerine algoritmanın eğitilecek örnekleri seçmesi, en zor ve en önemli olan örneklerin ise insana işaretlendirilmesi sağlanarak benzer ya da daha iyi sonuçlar elde edilebilir. Biz de çalışmamızda kolay ve zor örnekleri bulabilmek için aktif öğrenme tekniklerini değerlendirdik, çalışmamızda denediğimiz farklı sorgulama stratejilerinden aşağıdaki gibi bahsettik.

5.1.1 Rastgele örneklem seçimi (Random selection)

Etiketlenecek verilerin tamamen *rastgele seçildiği* yöntemdir. Aslında bu yönteme pasif öğrenme de diyebiliriz, diğer metotlarla karşılaştırdığımızda yüksek maliyetli olmasına rağmen veri setindeki dağılımı bize yansıtma olanağı olduğundan denenmesi faydalıdır. Bu formül denklem 5.1'de gösterilmiştir.

$$QS_{RandomSelection} = random.choice(x) \tag{5.1}$$

5.1.2 En az güven seçimi (Least confidence selection)

Modelin örnek için yaptığı tahminin olasılığı düşükse o zaman örnek ya zordur ya da model tarafından daha önce karşılaşılmamıştır. Bu verileri bulabilmek için *en az güven*

seçimi kullanılabilir. Bu yöntemde tüm veri noktaları için en yüksek olasılıklı örnekler alınır. Bu hesaplamalara ait denklemler 5.2 ve 5.3'te verilmiştir. Bu sonuçlar küçükten büyüğe doğru sıralanır, en küçükleri en az güvenilir olanlarıdır.

$$\hat{y} = \underset{y}{argmax}\, P(y \mid x) \tag{5.2}$$

$$QS_{LeastConfidence} = \underset{x}{argmax}\, P(1 - P(\hat{y} \mid x)) \tag{5.3}$$

5.1.3 En çok güven seçimi (Max confidence selection)

En az güven seçiminin tam tersi diyebiliriz, modelin eğitiminde doğrudan kullanmak için bize uygun olan örnekleri sağlayabilir. Benzer şekilde her bir veri noktasındaki en yüksek olasılıklı örnekler alınıp büyükten küçüğe sıralandığında en yüksek değerliler bize en kolay örnekleri verirler. Bu formüle ait denklem 5.4'teki gibidir.

$$QS_{MaxConfidence} = \underset{x}{argmax}\, P(P(\hat{y} \mid x)) \tag{5.4}$$

5.1.4 Marj belirsizliği seçimi (Margin selection)

Bu yöntem de zor örneklerin bulunmasına yardımcı olur, modelin kararsız kaldığı noktaları bularak işaretlenmesini sağlayabilir. Bunu da veri noktalarındaki en yüksek birinci ve ikinci olasılıkların farkını alarak yapar, formül olarak gösterimi denklem 5.5'teki gibidir. Bu fark yüksekse model için o seçimin kolay olduğu anlamına gelir, fark ne kadar az ise modelin kararsızlığı o kadar yüksektir. Ve bunların işaretleyici tarafından etiketlenmesinde fayda vardır.

$$QS_{MarginSelection} = \underset{x}{argmin}\, (P(\hat{y}_{max} \mid x) - P(\hat{y}_{max-1} \mid x)) \tag{5.5}$$

5.1.5 Entropi seçimi (Entropy selection)

Denediğimiz son yöntem ise *entropi seçimi* yöntemidir. Bu yöntemde de kararsızlık üzerinden ilerlenmektedir, bunun diğerlerinden farkı olası tüm sınıfların dağılımını hesaba katarak sonuç vermesidir. Yani model bir sonuçtan eminse belirli bir veri

noktasındaki bir sınıf için yapılan tahminin olasılığının yüksek, diğer sınıfların olasılığının ise az olduğu anlamına gelir. Bu sebeple sınıflandırma problemlerinde başarılı sonuçlar verebilmektedir. Bu yönteme ait genel formül denklem 5.6'daki gibi gösterilmiştir.

$$QS_{EntropySelection} = \underset{x}{argmax} \left(- \sum_{i} P(\hat{y}_i \mid x) \log P(\hat{y}_i \mid x) \right) \tag{5.6}$$

5.2 Zayıf Denetleme

Zayıf denetim, büyük miktarda eğitim verisini etiketlemek için gürültülü, sınırlı veya kesin olmayan kaynakların kullanıldığı bir makine öğrenimi dalıdır. Elle etiketlenmiş veri kümelerini oluşturmak çok maliyetli ve zaman alıcıdır. Bu yaklaşım sayesinde, bazı veriler ve fonksiyonlar kullanılarak etiketlenmemiş verilere etiket atanması sağlanır ve elle etiketlenmiş veri setleri oluşturma sürecindeki yük hafifletilir. Zayıf denetim kullanacak kişiler ya alanındaki uzman kişilerden yardım alarak ya da sezgisel kurallar ve bilgi tabanları gibi kaynakları kullanarak kod parçaları geliştirirler. Bunlar sayesinde etiketsiz veriler hızlı ve kolay bir şekilde etiketlenirler.

Biz de zayıf denetimin faydalarından yararlanmak için Snorkel [33] aracını kullandık ve fonksiyonlar oluşturduk. Bunları *olayKontrol* ve *etiketAta* fonksiyonu olarak adlandırdık.

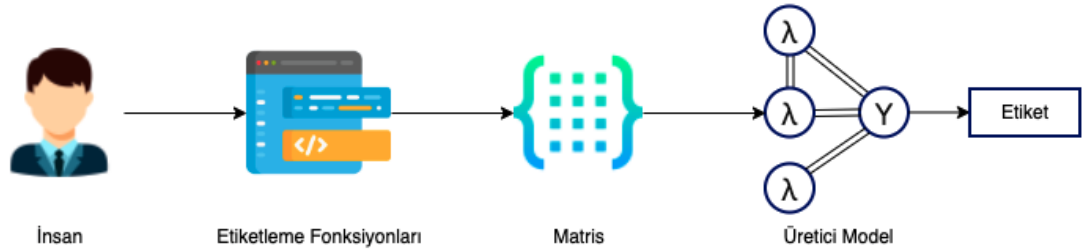
olayKontrol fonksiyonu önceden tanımladığımız olaylara ait muhtemel anahtar kelimeleri içeren bir liste alır ve cümlelerde bu anahtar kelimelerin olup olmadığını kontrol eder. Anahtar kelime listede bulunuyorsa bu cümle bir olay içeriyor deriz ve olay türlerinin ve argümanlarının belirlenmesi için *etiketAta* fonksiyonunu çağırarak çıktısını döneriz. Anahtar kelime olmadığı durumda ise cümlede olay bulunmadığından boş etiket atarız.

etiketAta fonksiyonu ise etiketlenmesi istenen metin için aldığı gürültülü sinyalleri kullanılarak nihai etiketleri döner. Buradaki nihai amacımız daha iyi bir tahmin elde edebilmek, bu sebeple bilgiyi tek bir kaynaktan almak yerine birden çok kaynağı birleştirilerek kullanıyoruz. Bunu gerçekleştirebilmek için önce Snorkel'in kullanacağı etiketleme fonksiyonları tanımladık.

Olay sınıflandırma için 2 fonksiyon tanımladık ve bu fonksiyonlarda şu sinyallerden yararlandık: olay sınıflandırma modelinin etiketleri, varlık ismi tanımlama etiketleri. Olay sınıflandırma modelinin etiketleri cümlelerdeki kelimelerin olayı tetikleyen bir ifade olup olmadığını varsa ne olduğunu belirtir. Kelimeye bir etiket atandıysa bu kelimenin tetikleyici ifade olduğunu söyleyebiliriz. Olayı ifade eden tetikleyiciler varlık ismi olamazlar, bu bilgiden yola çıkarak eğer bir kelime için varlık etiketi atandıysa o kelime tetikleyici olamaz deriz.

Argüman sınıflandırma için de 2 fonksiyon tanımladık ve şu sinyallerden yararlandık: olay sınıflandırma modelinin etiketleri, argüman sınıflandırma modelinin etiketleri. Argümanlar olaya dahil olan katılımcılardır, eğer bir cümlede olay yoksa argüman da yok diyebiliriz. Argüman sınıflandırma modelinin etiketleri kelimelerde bir argüman olup olmadığını, argüman varsa ne olduğunu ifade eder. Etiket atanan kelimeler argümanları ifade ederken diğerleri için bundan bahsedemeyiz.

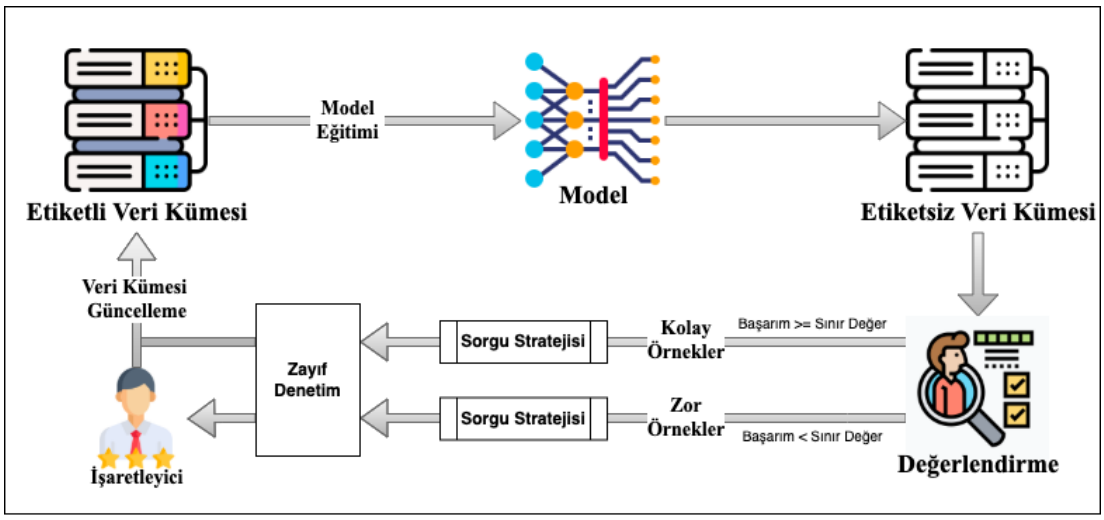
Snorkel aracı yazdığımız etiketleme fonksiyonlarını etiketsiz veriler üzerine uyguluyor ve sonucunda verilerden gelen etiketlerin yer aldığı etiket matrisini çıktı olarak veriyor. Sonrasında fonksiyonların doğruluğunu ve ilişkisini tahmin edebileceği bir üretici model(generative model) oluşturuyor. Nihayetinde bir dizi olasılıksal etiket çıkartıyor, biz de olasılığı eşik değerin üstündeki etiketi nihai etiket olarak değerlendiriyoruz. Anlattığımız süreç Şekil 5.1'de olduğu gibi özetlenebilir.



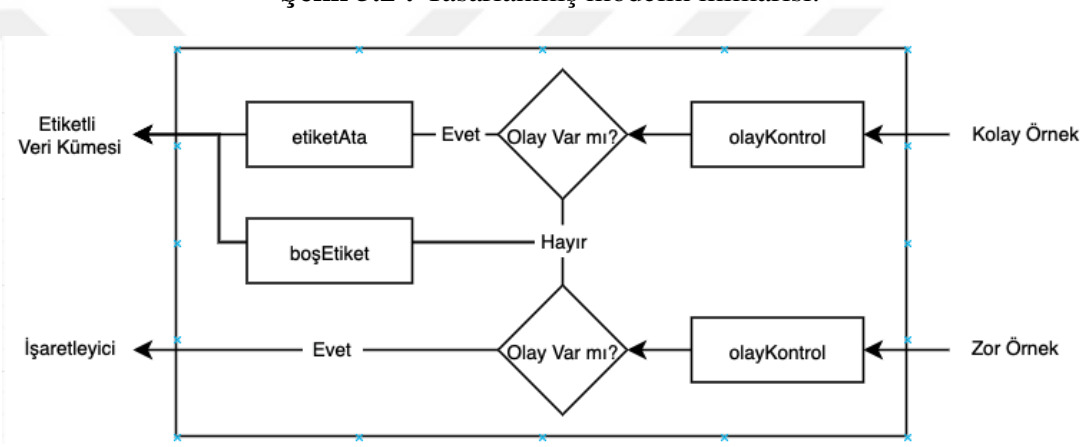
Şekil 5.1 : Zayıf denetim sürecinin mimarisi.

5.3 Zayıf Denetleme ile Geliştirilmiş Aktif Öğrenme Döngüsü

Önceki bölümlerde zayıf denetimin ve aktif öğrenme tekniklerinin güçlü yönlerini ve faydalarını gördük. Her ikisinden de faydalanabilmek için zayıf denetimi aktif



Şekil 5.2 : Tasarlanmış modelin mimarisi.



Şekil 5.3 : Zayıf denetim modülü.

öğrenme döngüsüne dahil ettik. Buradaki amacımız aktif öğrenme ile en önemli ve zor örnekleri seçip insanın işaretlemesini sağlamak ve kolay örneklerin ise zayıf denetim tarafından otomatik olarak etiketlenmesini sağlamak. Bunun için oluşturmuş olduğumuz mimari Şekil 5.2 ve 5.3'teki gibidir.

Yaşam döngüsü hayatına iki veri kümesi ile başlar, bunlardan birincisi etiketli verilerin olduğu küme, diğeri ise işaretlenmemiş verilerin bulunduğu kümedir. Etiketli veri kümesindeki veriler kullanılarak modelin eğitilmesi sağlanır, eğitilen model etiketsiz veri kümesindeki verilerin etiketleri hakkında tahminde bulunur. Yapılan tahminlerin olasılıklarına bakarak belirlemiş olduğumuz eşik değerin altındaki verileri zor örnekler olarak adlandırıp ayrı bir yoldan ilerlemesini sağlıyoruz, tam tersi durumda da

olasılığı yüksek çıkanları farklı bir yoldan ilerletiyoruz. Her iki yol için de farklı sorgu stratejileri belirledik, kolay örnekler için *en çok güven seçimi* sorgu stratejisini kullandık, bu seçilen örnekler zayıf denetim modülü tarafından atanan etiketlerle etiketlenmiş veri kümesine eklenecekler, kaliteyi düşürmemeleri için bunların arasında en güvenilir olanları seçiyoruz. Zor örnekler için de *en az güven seçimi* gibi sorgu stratejileri benimsedik buradaki amacımız en kararsız ve modelin performansı üzerinde en büyük etkiyi yaratacak örnekleri bulmak, bunların insan tarafından etiketlemesini sağlamak. Sorgu stratejileri tarafından seçilen örnekler zayıf denetim modülüne girdi olarak verilirler. Bu modülde kullanılmak üzere geliştirdiğimiz zayıf denetim fonksiyonları, örneklere doğrudan bir etiket atayabilir veya işaretlemesi için insana bırakabilir. Son aşamada ise etiketlendirilmiş örnekler etiketli veri kümesine dahil edilir ve etiketlenmemiş veri kalmayana kadar bu döngü devam eder.

Olay çıkarma görevini iki aşamada ele alıyoruz, ilk aşaması cümlelerdeki olayların sınıflandırılmasıdır, bunun için haber metinlerindeki olayı tetikleyen ifadeler belirlenir ve olayın türü tahmin edilmeye çalışılır. İkinci aşama ise olaya katılan argümanların belirlenmesidir, bu aşamada argümanların sınıflandırılması yapılır. Yaşam döngüsü iki aşamada için de benzerdir, tek fark zayıf denetim modülünde yer alan *etiketAta* fonksiyonundan kaynaklanmaktadır.

Zayıf denetim modülüne ait sözde kod algoritma 1 ve 2'deki gibidir. Algoritmalardaki önemli noktaları belirtecek olursak, *olayKontrol* fonksiyonu metinde bir olayın olup olmadığını tahmin eder. Eğer metinde olay bulunduysa *etiketAta* fonksiyonu sahip olduğu bilgileri ve tahminleri kullanarak ona son etiketlerini atar. Olayın olmadığı durumda ise tüm etiketler boş olarak işaretlenir. Kolay örnekler için akış bu iken, zor örneklerde *etiketAta*'ın yerini işaretleyici alır, ve insan tarafından işaretlenen etiketler kullanılır.

Algorithm 1 Olay Sınıflandırma

Girdi :
etiketlenmemiş veriler (Zor) UH,
etiketlenmemiş veriler (Kolay) UE,
varlık ismi tanımlama N
sınıflandırıcı f

Çıktı :
etiketler y

for $k \leftarrow 1$ to $length(UE)$ **do**
  $olayIsareti \leftarrow olayKontrol(UE_k)$
  if $olayIsareti$ is true **then**
    $y \leftarrow y + etiketAta(f, UE_k, N_k)$
  else
    $y \leftarrow y + bosEtiket$
  end if
end for
for $k \leftarrow 1$ to $length(UH)$ **do**
  $olayIsareti \leftarrow olayKontrol(UH_k)$
  if $olayIsareti$ is true **then**
    $y \leftarrow y + elleEtiketleme(UH_k)$
  else
    $y \leftarrow y + bosEtiket$
  end if
end for
return y

Algorithm 2 Argüman Sınıflandırma

Girdi :
etiketlenmemiş veriler (Zor) UH,
etiketlenmemiş veriler (Kolay) UE,
olay türü T,
sınıflandırıcı f

Çıktı :
etiketler y

for $k \leftarrow 1$ to $length(UE)$ **do**
  $olayIareti \leftarrow olayKontrol(UE_k)$
  if $olayIsareti$ is true **then**
    $y \leftarrow y + etiketAta(f, UE_k, T_k)$
  else
    $y \leftarrow y + bosEtiket$
  end if
end for
for $k \leftarrow 1$ to $length(UH)$ **do**
  $olayIsareti \leftarrow olayKontrol(UH_k)$
  if $olayIsareti$ is true **then**
    $y \leftarrow y + elleEtiketleme(UH_k)$
  else
    $y \leftarrow y + bosEtiket$
  end if
end for
return y

6. DEĞERLENDİRMELER

6.1 Olay ve Argüman Çıkarımı

Bu bölümde, denediğimiz yöntemlerin uygulanabilirliğini araştırdık, modellerin çıktılarını analiz ettik ve son olarak yöntemlerin daha fazla etiketli veri elde etmede ve insan çabasını azaltmada ne kadar etkili olduğunu göstermek istedik.

Öncelikle karşılaştırmalarımızı yapabilmek ve ileriki çalışmalara rehberlik edebilmek için bir temel oluşturduk. Bunun için elle etiketlemiş olduğumuz veri kümemizi %80 ve %20 oranında eğitim ve test kümesi olmak üzere ikiye böldük. Ardından önceden eğitilmiş modelleri [43] araştırdık. Modelleri sıfırdan eğitmek yerine veri kümemizdeki eğitim verilerini kullanarak ayarlama(fine-tuning) yaptık. Olay çıkarımı çalışması token sınıflandırma problemi olarak ele alındığından dolayı biz de token düzeyinde tahmin yapmamızı sağlayan modelleri seçtik.

Çalışmada kullandığımız modellerin parametreleri şu şekildedir: Öğrenme oranı(Learning Rate) 2e-5, sekans uzunluğu(Sequence Length) 128, eğitim aşamasındaki grup boyutu(Batch Size per Device During Training) 3, değerlendirme aşamasındaki grup boyutu(Batch Size for Evaluation) 3, toplam eğitim devri(Total Number of Training Epochs) 1, ağırlık düşüşü gücü (Strength of Weight Decay) 0.01.

Çizelge 6.1 : Temel model sonuçları.

Model	**F1 Score**	
	Olay sınıflandırıcı	**Argüman sınıflandırıcı**
BERT	85,89	65,81
CONVBERT	80,63	61,86
DISTILBERT	78,57	58,16
ELECTRA	80,33	60,28
CRF	65,82	49,22

Kullanılan modeller ve makro ortalamalı F1 puanları Çizelge 6.1'de görülebilir.

Çizelge 6.2 : Olay sınıflandırma sonuçları.

Yöntemler	Eğitim		Test	F1 Skoru	Oransal Değişim	Kurtarılan İnsan Eforu
	Toplam Cümle Sayısı	*Elle Etiketlenen Cümle Sayısı*	*Toplam Cümle Sayısı*			
Referans	5K	5K	1K	85,89		
AL	3,75K	3,75K	1K	85,57	-0,32%	25%
AL + WS	5K	2,3K	1K	86,51	+0,7%	54%
WS	322K	5K	1K	88,39	+2,91%	98%

Çizelge 6.3 : Argüman sınıflandırma sonuçları.

Yöntemler	Eğitim		Test	F1 Skoru	Oransal Değişim	Kurtarılan İnsan Eforu
	Toplam Cümle Sayısı	*Elle Etiketlenen Cümle Sayısı*	*Toplam Cümle Sayısı*			
Referans	5K	5K	1K	65,81		
AL	3,5K	3,5K	1K	65,77	-0,06%	30%
AL + WS	5K	2K	1K	67,34	+2,32%	60%
WS	322K	5K	1K	74,87	+13,76%	98%

Sonuçları incelediğimizde her iki sınıflandırma görevi için de BERT modelinin en iyi sonuçları verdiğini görüyoruz. Bu sonuçlar olay sınıflandırması için %85,89 ve argüman sınıflandırması için %65,81 olarak kaydedilmiştir. Bu sınıflandırmalar yapılırken, sınıflar her zaman eşit dağılmadığından her bir sınıf için kesinlik(precision), hassasiyet(recall) ve f1-skoru(f1-score) ayrı ayrı hesaplanmış argüman sonuçları Çizelge A.1'de, olay sınıflandıra sonuçları ise Çizelge A.2, A.3 ve A.4'te gösterilmiştir.

Denediğimiz metotlara ait sonuçlar, olay sınıflandırması için Çizelge 6.2'de ve argüman sınıflandırması için Çizelge 6.3'te gösterilmiştir. Çizelgelerdeki ilk sütun [Yöntemler] hangi yöntemi kullandığımızı gösterir. İkinci sütunda [Toplam Cümle Sayısı], model eğitiminde kullanılan toplam veri miktarı gösterilirken, üçüncü sütunda [Elle Etiketlenen Cümle Sayısı] toplam veri sayısı içinde insan tarafından etiketlenen veri miktarı gösterilir. Çalışmalarımızı cümle düzeyinde yaptığımız için bu alanlarda cümle sayılarını paylaşıyoruz. Dördüncü sütun [Toplam Cümle Sayısı], eğitilen modelin test edileceği veri kümesindeki veri sayısını gösterir. Beşinci sütun [F1 Puanı], yöntemlerin makro ortalamalı F1 puanlarını içerir. Altıncı sütunda [Oransal Değişim], temel yönteme kıyasla bu yöntemin yüzde kaçlık bir değişiklik sağladığını bize gösterir. Son sütunda [Kurtarılan İnsan Eforu], yukarıda belirtilen yöntemler denendiğinde insan eforunun yüzde kaçının kazanıldığı gösterilmektedir.

Sonuçları yorumlamak için, referans veri kümesindeki tüm etiketli verileri kullanarak model eğittik ve sonuçlarını temel aldık. Çizelgelerin ilk satırında bu modelin sonuçlarına ait bilgiler yer almaktadır.

Aktif öğrenme, zayıf denetimde olduğu gibi bize etiketli veriler üretmez, bu yüzden sadece aktif öğrenmeyi kullanmamızdaki amaç daha az sayıda etiketli veri kullanarak temel performans sonucuna erişmektir. Deneylere baktığımızda olay sınıflandırma için 3.75K veri ve argüman sınıflandırması için 3.5K veri kullandığımızda 5K veri ile eğittiğimiz temel modelin performansına çok küçük farkla ulaşabileceğimizi görüyoruz. Sonuçlar için çizelgelerin ikinci satırlarına bakılabilir.

Aktif öğrenme ve zayıf denetimi birlikte uyguladığımızda sadece 2K veri elle etiketlenmiş olarak kullanılır ve kalan 3K veri otomatik olarak etiketlenir. Aktif

öğrenmeyi ve zayıf denetimi bir arada kullanmak, yalnızca ortaya koyulan efordan tasarruf sağlamakla kalmaz, aynı zamanda performansı da artırır. Bu da zayıf denetimin kaliteli çıktılar ürettiğini kanıtlar. Aynı çıkarımı argüman sınıflandırması için de yapabiliriz. Çizelgelerin üçüncü satırlarında detaylandırılmıştır.

Zayıf denetimden elde ettiğimiz iyi sonuçlarına dayanarak, haberlerden çıkardığımız etiketlenmemiş havuz verilerini etiketledik. Ardından model eğitimi için elle etiketlenen 5K veriyi ve zayıf denetim tarafından işaretlenen 317K veriyi kullandık. Sonuç olarak, olay sınıflandırma için %2,91 artış, argüman sınıflandırma için %13,76 artış görüyoruz. Bu bilgiler için çizelgelerin son satırlarına bakılabilir.

6.2 Veri Kümesi

Bu bölümde elle işaretlediğimiz veri kümesi ile, zayıf denetim kullanarak otomatik olarak etiketlenmesini sağladığımız veri kümesini bu alanda yapılan önceki diğer önemli çalışmalarla karşılaştırıyoruz. Bu istatistiksel bilgiler Çizelge 6.4'te görülebilmektedir. Karşılaştırılan alanlar; işaretlenen belge sayıları, tüm dokümanlardaki toplam cümle ve kelime sayıları, bunun yanında ayrıca veri kümesindeki olayları ifade eden tetikleyici sayıları ve olaya katılan argüman sayılarıdır. Ek olarak veri kümelerinde yer alan tür ve alt tür sayılarını da gösteriyoruz.

Çizelge 6.4 : Veri kümelerine ait istatistiksel bilgiler.

	ACE05	SENTIVENT	Referans Veri Kümesi	Otomatik Etiketlenen Veri Kümesi
Dokümanlar	599	288	600	34.146
Cümleler	18.927	6.883	6.249	317.696
Kelimeler	303.000	170.398	117.206	6.063.037
Olaylar	5.055	6.203	7.824	303.480
Argümanlar	6.040	13.675	9.148	377.011
Tür	8	18	25	25
Alt Tür	33	64	104	104

Çizelgedeki değerleri okuduğumuzda belge sayısı olarak referans veri kümemizin ACE05 veri kümesiyle benzer sayıda doküman içerdiğini, SENTIVENT veri kümesinden ise daha fazla belgeye sahip olduğunu görüyoruz. Veri kümelerindeki cümle ve kelime sayılarına baktığımızda ise diğer derlemlerin bu alanda daha

fazla metin içerdiği görülebilir. Bu sayılardan yola çıkarak, incelenen diğer veri kümelerindeki dokümanların Türkçe haber makalelerinden aldığımız metinlerden uzun olup daha fazla cümle ve kelime içerdiği söylenebilir. Derlemlerdeki olay sayılarına göz attığımızda çalışmamızın diğer çalışmalardan daha fazla olay sayısına sahip olduğunu görmekteyiz, bunun sebebi oluşturmuş olduğumuz olay kapsamının diğerlerinden daha geniş ve detaylı olmasıdır. Argümanlar konusunda ACE05 ile çalışmamızı karşılaştıracak olursak daha fazla olay içerdiğinden argüman sayımızın da daha fazla olması tutarlılık gösteriyor, SENTIVENT'i değerlendirdiğimizde ise onların olayları daha fazla argüman kullanarak açıkladığını görüyoruz. Genel olarak baktığımızda elle oluşturduğumuz veri kümemizin bu alanda yapılan en önemli çalışmalarla rekabet edebildiğini söyleyebiliyoruz.

Son olarak, otomatik olarak elde ettiğimiz veri kümesine bakacak olursak, derlemimizin elle işaretlenen diğer veri kümelerine kıyasla her alanda daha büyük boyuta sahip olduğunu rahatlıkla söyleyebiliriz.

7. SONUÇLAR

Metinden olay çıkarma, doğal dil işleme dünyasında önemli bir görevdir ve gerçek dünyadaki değişiklikleri yakalamak, bunları olay türüne göre sınıflandırmak ve olay argümanlarını belirlemek için kullanılır. Finansal ve ekonomik alanda olay çıkarma derlemleri çeşitli dillerde mevcuttur ancak bildiğimiz kadarıyla Türkçe'de herhangi bir derlem bulunmamaktadır.

Bu çalışmayla, ekonomik ve finansal metin madenciliği alanında eğitim verisi olarak kullanılabilecek, olaylar hakkında zengin bir kaynak sağlayan iyi tanımlanmış, yüksek kaliteli bir Türkçe veri kümesi sunuyoruz. Bununla birlikte ileride yapılacak olan Türkçe olay çıkarımı çalışmaları için de temel sağlıyoruz.

Veri miktarı zor problemleri anlamada ve yüksek başarıma sahip modeller oluşturmada oldukça önemli bir yer tutar, biz de kaliteyi düşürmeden insan açıklama çabasını azaltmak ve veri kümesini genişletmek için aktif öğrenme ve zayıf denetim yöntemlerini araştırdık, uyguladık ve ilk sonuçlarını kaydettik. Yaptığımız çalışmalar sonucunda zayıf denetim yöntemlerini kullanarak veri kümesindeki yüzlerce etiketli belgeyi on binlerce etiketli belgeye çıkardık. Deneysel sonuçlarımız, uygulanan yöntemin yalnızca yüksek kaliteli etiketli veriler ürettiğini değil, aynı zamanda model performansını da arttırdığını göstermektedir.

Gerçekleştirdiğimiz çalışmanın gelecekteki Türkçe olay çıkarım araştırmalarında yüksek performanslı model elde etmede veya veriye aç görevlerde eğitim verilerini arttırmada kullanılabileceğine inanıyoruz.

Hazırladığımız derlem https://github.com/kadirsinas/TFEEC linki aracılığı ile herkese açık olarak sunulmuştur.

KAYNAKLAR

[1] **Eaton, J.**, **Gaubitch, N.D.**, **Moore, A.H. ve Naylor, P.A.** (2015). The ACE challenge—Corpus description and performance evaluation, *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, IEEE, s.1–5.

[2] **Chen, C. ve Ng, V.** (2012). Joint modeling for chinese event extraction with rich linguistic features, *Proceedings of COLING 2012*, s.529–544.

[3] **Riloff, E.** *ve diğerleri* (1993). Automatically constructing a dictionary for information extraction tasks, *AAAI*, cilt 1, Citeseer, s.2–1.

[4] **Borsje, J.**, **Hogenboom, F. ve Frasincar, F.** (2010). Semi-automatic financial events discovery based on lexico-semantic patterns, *International Journal of Web Engineering and Technology*, *6*(2), 115–140.

[5] **Nguyen, T.H.**, **Cho, K. ve Grishman, R.** (2016). Joint event extraction via recurrent neural networks, *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, s.300–309.

[6] **Nguyen, T.H. ve Grishman, R.** (2015). Event detection and domain adaptation with convolutional neural networks, *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, s.365–371.

[7] **Arendarenko, E. ve Kakkonen, T.** (2012). Ontology-based information and event extraction for business intelligence, *International Conference on Artificial Intelligence: Methodology, Systems, and Applications*, Springer, s.89–102.

[8] **Aguilar, J.**, **Beller, C.**, **McNamee, P.**, **Van Durme, B.**, **Strassel, S.**, **Song, Z. ve Ellis, J.** (2014). A comparison of the events and relations across ACE, ERE, TAC-KBP, and FrameNet annotation standards, *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, s.45–53.

[9] **Ahn, D.** (2006). The stages of event extraction, *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, s.1–8.

[10] **Chieu, H.L. ve Ng, H.T.** (2002). A maximum entropy approach to information extraction from semi-structured and free text, *Aaai/iaai*, *2002*, 786–791.

[11] **Nguyen, T. ve Grishman, R.** (2018). Graph convolutional networks with argument-aware pooling for event detection, *Proceedings of the AAAI Conference on Artificial Intelligence*, cilt 32.

[12] **Jacobs, G. ve Hoste, V.** (2021). SENTiVENT: enabling supervised information extraction of company-specific events in economic and financial news, *Language Resources and Evaluation*, 1–33.

[13] **Mintz, M.**, **Bills, S.**, **Snow, R. ve Jurafsky, D.** (2009). Distant supervision for relation extraction without labeled data, *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, s.1003–1011.

[14] **Chen, Y.**, **Liu, S.**, **Zhang, X.**, **Liu, K. ve Zhao, J.** (2017). Automatically labeled data generation for large scale event extraction, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, s.409–419.

[15] **Zeng, Y.**, **Feng, Y.**, **Ma, R.**, **Wang, Z.**, **Yan, R.**, **Shi, C. ve Zhao, D.** (2018). Scale up event extraction learning via automatic training data generation, *Proceedings of the AAAI Conference on Artificial Intelligence*, cilt 32.

[16] **Abney, S.** (2002). Bootstrapping, *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, s.360–367.

[17] **Huang, R. ve Riloff, E.** (2012). Bootstrapped training of event extraction classifiers, *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, s.286–295.

[18] **Hachey, B.**, **Alex, B. ve Becker, M.** (2005). Investigating the effects of selective sampling on the annotation task, *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, s.144–151.

[19] **Hoi, S.C.**, **Jin, R. ve Lyu, M.R.** (2006). Large-scale text categorization by batch mode active learning, *Proceedings of the 15th international conference on World Wide Web*, s.633–642.

[20] **Ringger, E.**, **McClanahan, P.**, **Haertel, R.**, **Busby, G.**, **Carmen, M.**, **Carroll, J.**, **Seppi, K. ve Lonsdale, D.** (2007). Active learning for part-of-speech tagging: Accelerating corpus annotation, *Proceedings of the Linguistic Annotation Workshop*, s.101–108.

[21] **Reichart, R. ve Rappoport, A.** (2007). Self-training for enhancement and domain adaptation of statistical parsers trained on small datasets, *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, s.616–623.

[22] **Zhu, J. ve Hovy, E.** (2007). Active learning for word sense disambiguation with methods for addressing the class imbalance problem, *Proceedings*

of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), s.783–790.

[23] **Liao, S. ve Grishman, R.** (2011). Using prediction from sentential scope to build a pseudo co-testing learner for event extraction, *Proceedings of 5th International Joint Conference on Natural Language Processing*, s.714–722.

[24] **Nguyen, T.H.**, **Fu, L.**, **Cho, K. ve Grishman, R.** (2016). A two-stage approach for extending event detection to new types via neural networks, *Proceedings of the 1st Workshop on Representation Learning for NLP*, s.158–165.

[25] **Wang, X.**, **Zhang, Y.**, **Li, Q.**, **Ren, X.**, **Shang, J. ve Han, J.** (2019). Distantly supervised biomedical named entity recognition with dictionary expansion, *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, s.496–503.

[26] **Huang, L.**, **Ji, H.**, **Cho, K. ve Voss, C.R.** (2017). Zero-shot transfer learning for event extraction, *arXiv preprint arXiv:1707.01066*.

[27] **Karamanolakis, G.**, **Mukherjee, S.**, **Zheng, G. ve Awadallah, A.H.** (2021). Self-training with weak supervision, *arXiv preprint arXiv:2104.05514*.

[28] **Lison, P.**, **Hubin, A.**, **Barnes, J. ve Touileb, S.** (2020). Named entity recognition without labelled data: A weak supervision approach, *arXiv preprint arXiv:2004.14723*.

[29] **Stewart, R. ve Ermon, S.** (2017). Label-free supervision of neural networks with physics and domain knowledge, *Thirty-First AAAI Conference on Artificial Intelligence*.

[30] **Druck, G.**, **Settles, B. ve McCallum, A.** (2009). Active learning by labeling features, *Proceedings of the 2009 conference on Empirical methods in natural language processing*, s.81–90.

[31] **Reschke, K.**, **Jankowiak, M.**, **Surdeanu, M.**, **Manning, C.D. ve Jurafsky, D.** (2014). Event extraction using distant supervision, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, s.4527–4531.

[32] **Rao, S.**, **Marcu, D.**, **Knight, K. ve Daumé III, H.** (2017). Biomedical event extraction using abstract meaning representation, *BioNLP 2017*, s.126–135.

[33] **Ratner, A.**, **Bach, S.H.**, **Ehrenberg, H.**, **Fries, J.**, **Wu, S. ve Ré, C.** (2017). Snorkel: Rapid training data creation with weak supervision, *Proceedings of the VLDB Endowment. International Conference on Very Large Data Bases*, cilt 11, NIH Public Access, s.269.

[34] **Uzun, A. ve Tantuğ, A.C.** (2021). ITUTime: Turkish Temporal Expression Extraction and Normalization, *International Symposium on Distributed Computing and Artificial Intelligence*, Springer, s.74–85.

[35] **Yimam, S.M.**, **Gurevych, I.**, **de Castilho, R.E. ve Biemann, C.** (2013). Webanno: A flexible, web-based and visually supported system for distributed annotations, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, s.1–6.

[36] **Güneş, A. ve TantuĞ, A.C.** (2018). Turkish named entity recognition with deep learning, *2018 26th Signal Processing and Communications Applications Conference (SIU)*, s.1–4.

[37] **Ferguson, J.**, **Lockard, C.**, **Weld, D.S. ve Hajishirzi, H.** (2018). Semi-supervised event extraction with paraphrase clusters, *arXiv preprint arXiv:1808.08622*.

[38] **Wang, X.**, **Han, X.**, **Liu, Z.**, **Sun, M. ve Li, P.** (2019). Adversarial training for weakly supervised event detection, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, s.998–1008.

[39] **Rusu, D.**, **Hodson, J. ve Kimball, A.** (2014). Unsupervised techniques for extracting and clustering complex events in news, *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, s.26–34.

[40] **Chun, H.W.**, **Hwang, Y.S. ve Rim, H.C.** (2004). Unsupervised event extraction from biomedical literature using co-occurrence information and basic patterns, *International Conference on Natural Language Processing*, Springer, s.777–786.

[41] **Chambers, N. ve Jurafsky, D.** (2011). Template-based information extraction without the templates, *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, s.976–986.

[42] **Marşan, B.**, **Kara, N.**, **Özçelik, M.**, **Arıcan, B.N.**, **Cesur, N.**, **Kuzgun, A.**, **Sanıyar, E.**, **Kuyrukçu, O. ve Yıldız, O.T.** (2021). Building the turkish framenet, *Proceedings of the 11th Global Wordnet Conference*, s.118–125.

[43] **Schweter, S.** (2020). *BERTurk - BERT models for Turkish*, `https://doi.org/10.5281/zenodo.3770924`.

EKLER

EK A : Olay Türü ve Argüman Sınıflandırma Sonuçları

EK A : Olay Türü ve Argüman Sınıflandırma Sonuçları

Çizelge A.1 : Argüman sınıflandırma modelinin sınıf bazındaki sonuçları.

Sınıf (Class)	Kesinlik (Precision)	Hassasiyet (Recall)	F1-Skor (F1-Score)
Analist	0.50	0.67	0.57
Kişi	0.61	0.84	0.71
Oran	0.90	0.97	0.93
Para	0.95	0.96	0.96
Proje	0.57	0.40	0.47
Sayı	0.68	0.81	0.74
Tahmin	0.00	0.00	0.00
Tavsiye	0.62	1.00	0.77
Unvan	0.45	0.65	0.53
Yer	0.62	0.93	0.75
Zaman	0.80	0.88	0.84
Ürün	0.50	0.83	0.62
İş Alanı	0.57	0.89	0.70
Şirket	0.78	0.89	0.83

Çizelge A.2 : Olay sınıflandırma modelinin sınıf bazındaki sonuçları.

Sınıf (Class)	Kesinlik (Precision)	Hassasiyet (Recall)	F1-Skor (F1-Score)
Analist-Beklenti	0.80	1.00	0.89
Analist-Raporlama	1.00	1.00	1.00
Analist-Tavsiye	0.90	1.00	0.95
Anlaşma	0.87	1.00	0.93
Bilgi Değiştirme-Logo Değiştirme	0.75	1.00	0.86
Bilgi Değiştirme-İsim Değiştirme	1.00	1.00	1.00
Birleşme	0.71	1.00	0.83
Borsa-Halka Arz	1.00	1.00	1.00
Borsa-Pazara Girme	1.00	1.00	1.00
Borsa-Pazardan Çıkma	1.00	1.00	1.00
Borsa-İşlem Görme	1.00	1.00	1.00
Borç	1.00	1.00	1.00
Finansal Sonuç-Dönem Bilançosu	0.93	1.00	0.96
Finansal Sonuç-Kredi Notu	1.00	1.00	1.00
Finansal Sonuç-Mali Tablo	1.00	1.00	1.00
Finansal Sonuç-Satış Hacmi	0.95	1.00	0.97
Gelir-Artış	0.96	0.94	0.95
Gelir-Azalış	0.87	1.00	0.93
Gelir-Duyuru	0.78	0.90	0.84
Gider-Artış	0.00	0.00	0.00
Gider-Duyuru	0.40	0.50	0.44
Hisse Pay-Alış	0.97	0.85	0.90
Hisse Pay-Devri	0.95	1.00	0.97
Hisse Pay-Dönüşüm	1.00	1.00	1.00
Hisse Pay-Fiyat Hedefi	0.86	1.00	0.92
Hisse Pay-Geri Alma	1.00	1.00	1.00
Hisse Pay-Rehin	1.00	1.00	1.00
Hisse Pay-Satış	0.83	0.95	0.89
Kar-Artış	1.00	0.85	0.92
Kar-Azalış	1.00	1.00	1.00
Kar-Duyuru	0.89	1.00	0.94
Kurum-Açma	0.65	1.00	0.79
Kurum-Bölünme	1.00	1.00	1.00
Kurum-Kapama	0.67	0.80	0.73
Kurum-Yeniden Yapılanma	1.00	1.00	1.00
Kurum-İflas	1.00	1.00	1.00

Çizelge A.3 : Olay sınıflandırma modelinin sınıf bazındaki sonuçları (devam).

Sınıf (Class)	Kesinlik (Precision)	Hassasiyet (Recall)	F1-Skor (F1-Score)
Macroekonomi-Devalüasyon	1.00	1.00	1.00
Macroekonomi-Endeks	1.00	1.00	1.00
Macroekonomi-Enflasyon	0.94	1.00	0.97
Macroekonomi-Faiz	1.00	1.00	1.00
Macroekonomi-Milli Gelir	0.73	1.00	0.84
Macroekonomi-Pazar Payı	1.00	1.00	1.00
Macroekonomi-Zam	0.88	1.00	0.93
Macroekonomi-İşsizlik	1.00	1.00	1.00
Marka-Etik	0.80	1.00	0.89
Marka-Hibe	1.00	1.00	1.00
Marka-Kadına Şiddet	1.00	1.00	1.00
Marka-Reklam	0.92	1.00	0.96
Marka-Rüşvet	0.75	1.00	0.86
Marka-Skandal	0.67	1.00	0.80
Marka-Sosyal Sorumluluk	1.00	1.00	1.00
Marka-Spekülasyon	1.00	1.00	1.00
Marka-Sponsor	1.00	1.00	1.00
Marka-Yolsuzluk	1.00	1.00	1.00
Marka-Ödül	0.88	1.00	0.93
Marka-Üyelik	0.90	1.00	0.95
Marka-İmaj	0.00	0.00	0.00
Proje	1.00	1.00	1.00
Sermaye-Arttırma	0.83	0.93	0.88
Sermaye-Azaltma	1.00	1.00	1.00
Sermaye-Payı	0.25	1.00	0.40
Sermaye-Tavanı	1.00	1.00	1.00
Temettü	1.00	1.00	1.00
Yasa-Beraat	1.00	1.00	1.00
Yasa-Ceza	1.00	1.00	1.00
Yasa-Dava	1.00	1.00	1.00
Yasa-Soruşturma	0.80	1.00	0.89
Yasa-Suç Duyurusu	1.00	1.00	1.00
Yasa-Tazminat	1.00	1.00	1.00
Yasa-Tedbir	1.00	1.00	1.00
Yasa-Uzlaşma	1.00	1.00	1.00
Yasa-Yargılama	0.00	0.00	0.00
Yasa-Yasal İşlem	1.00	1.00	1.00
Yasa-İhlal	1.00	1.00	1.00
Yasa-İtiraz	1.00	1.00	1.00
Yatırım	0.99	1.00	0.99
Zarar-Artış	1.00	0.50	0.67
Zarar-Azalış	0.80	1.00	0.89
Zarar-Duyuru	0.94	0.97	0.95

Çizelge A.4 : Olay sınıflandırma modelinin sınıf bazındaki sonuçları (devam).

Sınıf (Class)	Kesinlik (Precision)	Hassasiyet (Recall)	F1-Skor (F1-Score)
Çalışan-Atama	0.75	1.00	0.86
Çalışan-Değişiklik	0.80	1.00	0.89
Çalışan-Grev	0.88	1.00	0.93
Çalışan-Kaza	1.00	1.00	1.00
Çalışan-Ölüm	0.33	1.00	0.50
Çalışan-İstifa	1.00	1.00	1.00
Çalışan-İstihdam	1.00	1.00	1.00
Çalışan-İş Sözleşmesi	1.00	0.98	0.99
Çalışan-İşten Çıkarma	1.00	1.00	1.00
Ürün-Alma	0.78	1.00	0.88
Ürün-Lansman	1.00	1.00	1.00
Ürün-Satma	0.83	0.85	0.84
Ürün-Çeşitlendirme	0.67	1.00	0.80
Ürün-Üretime Ara Verme	1.00	1.00	1.00
Ürün-Üretime Başlama	1.00	1.00	1.00
Ürün-Üretime Devam Etme	1.00	1.00	1.00
Ürün-Üretime Son Verme	1.00	1.00	1.00
Ürün-İmalat	0.98	1.00	0.99
İhale	1.00	1.00	1.00
İhracat	1.00	1.00	1.00
İş Alanı-Alış	0.25	1.00	0.40
İş Alanı-Açma	0.77	0.91	0.83
İş Alanı-Kapama	0.67	1.00	0.80
İş Alanı-Satış	0.82	0.82	0.82

ÖZGEÇMİŞ

Adı SOYADI: Kadir Şinas KAYNAK

ÖĞRENİM DURUMU:

- **Y. Lisans:** 2022, İstanbul Teknik Üniversitesi, Bilgisayar Mühendisligi Anabilim Dalı, Bilgisayar Mühendisliği Programı
- **Lisans:** 2017, Dokuz Eylül Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü

MESLEKİ DENEYİMLER VE ÖDÜLLER:

- Yazılım Geliştirme Uzmanı, 2017 - halen, GarantiBBVA

YÜKSEK LİSANS TEZİNDEN TÜRETİLEN YAYINLAR:

- **Kaynak K.**, Tantug A. (2022). *19th International Conference on Distributed Computing and Artificial Intelligence*, L'Aquila, Italy.