



## Patent classification with pre-trained Bert model

Selen Yücesoy Kahraman<sup>1\*</sup>, Alptekin Durmuşoğlu<sup>2</sup>, Türkey Dereli<sup>3</sup>

<sup>1</sup>Department of Industrial Engineering, Faculty of Engineering, Gaziantep University, 27310, Şehitkamil, Gaziantep, Türkiye

<sup>2</sup>Department of Industrial Engineering, Faculty of Engineering and Natural Science, Samsun University, 55420 Ondokuzmayıs, Samsun, Türkiye

<sup>3</sup>Office of President, Hasan Kalyoncu University, 27310, Şahinbey Gaziantep, Türkiye

### Highlights:

- Automated patent classification with BERT
- Selection of the most suitable model with hyperparameter analysis
- Compete with the results in the literature

### Keywords:

- Patent classification
- Bert,
- deep learning,
- text classification
- hyperparameter test

### Article Info:

Research Article

Received: 04.05.2023

Accepted: 05.03.2024

### DOI:

10.17341/gazimmfd.1292543

### Correspondence:

Author: Selen Yücesoy

Kahraman

e-mail:

yucesoy@gantep.edu.tr

phone: +90 342 317 2626

### Graphical/Tabular Abstract

Using an automatic classification system, patents assigned to the technical class to which they belong can pave the way for researchers and provide an environment in which they can create new inventions. This study presents an automatic patent classification analysis using the BERT algorithm. We attempted to increase prediction accuracy using hyperparameter analyses. The obtained results were at a level that competed with those in the literature. An accuracy of 58% was achieved at the subclass level.

Table A contains a summary of 9 different Bert analyzes performed in this study.

**Table A** Application of pre-trained BERT on Wipo alpha patent data

| Input Type            | Precision | Recall | F1   | Accuracy |
|-----------------------|-----------|--------|------|----------|
| Title                 | 0.47      | 0.48   | 0.46 | 0.49     |
| Abstract              | 0.54      | 0.54   | 0.53 | 0.54     |
| Claim                 | 0.55      | 0.54   | 0.53 | 0.54     |
| Description           | 0.57      | 0.57   | 0.56 | 0.57     |
| FulText               | 0.58      | 0.58   | 0.56 | 0.58     |
| Title+Abstract        | 0.55      | 0.56   | 0.54 | 0.56     |
| Abstract+Claim        | 0.55      | 0.55   | 0.54 | 0.55     |
| Title+Claim           | 0.57      | 0.56   | 0.55 | 0.56     |
| Title+ Abstract+Claim | 0.55      | 0.55   | 0.54 | 0.55     |

Learning Rate=2e-05, Batch Size=64, Epoch=5

### Purpose:

The aim of this study is to increase the performance of the automatic patent classification system used by patent researchers and/or patent experts when searching for patents. The most important problem of automatic patent classification is the misclassification of patents. This inaccuracy can create various problems, such as invalidating the patents in the courts or preventing the implementation of the patents. In order to avoid such problems, errors in automatic patent classification systems should be minimized.

### Theory and Methods:

In this study, the patent classification process was performed using the pre-trained Bert-base model. Focused on the Wipo alpha patent dataset. The different parts of patent texts in this data set were used. Moreover, different combination of this texts was also used as an input of classification problem. Hyperparametric analysis were also carried out.

### Results:

This study shows that the classification made with the Bert model can compete with the classification performances performed for the same data set in the literature. Analyzes were made with variations of hyperparameters such as learning rate, different input type, batch size and epoch. The best classification performance (Accuracy= 0.58) in this study was achieved when the description was taken as input, the learning rate was 2e-05, the batch size was 64 and the epoch value was 5.

### Conclusion:

In the study, classification performance was tried to be measured with different hyperparametric variables as well as different parts of patent texts. The result reached is at a level that can compete with the results obtained in this data set available in the literature. It is desirable to experiment with the classification process with BERT-large and other BERT family members in the future. Although it is not necessary to use word embedding in studies on BERT, in some studies word embedding can be used by ignoring this situation. In the future, it is desired to perform classification with a specific word embedding derived from patent texts.



## Ön eğitilmiş Bert modeli ile patent sınıflandırılması

Selen Yücesoy Kahraman<sup>1\*</sup>, Alptekin Durmuşoğlu<sup>2</sup>, Türkay Dereli<sup>3</sup>

<sup>1</sup>Gaziantep Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, 27310, Şehitkamil, Gaziantep, Türkiye

<sup>2</sup>Samsun Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Endüstri Mühendisliği Bölümü, 55420, Ondokuzmayıs, Samsun, Türkiye

<sup>3</sup>Hasan Kalyoncu Üniversitesi, Rektörlük, Şahinbey Gaziantep, Türkiye

### ÖNEÇİKANLAR

- Bu çalışmada otomatik patent sınıflandırma problemine Bert modeli ile çözüm aranmıştır
- Yapılan hiperparametre analizleri ile en uygun parametre kombinasyonu bulunmaya çalışılır
- Çalışmadan elde edilen sonuç literatürdeki çalışmalar ile rekabet eder düzeydedir

### Makale Bilgileri

Araştırma Makalesi

Geliş: 04.05.2023

Kabul: 05.03.2024

### DOI:

10.17341/gazimmfd.1292543

### Anahtar Kelimeler:

Patent sınıflandırması,  
Bert,  
derin öğrenme,  
metin sınıflandırması

### ÖZ

Patentler, bilgi teknolojilerindeki yeniliklerin korunmasına yardımcı olan ve bu yeniliklerin yaratıcısına belirli bir süre boyunca özel haklar sağlayan belgelerdir. Bu haklar, patent sahibine yeniliği ticari olarak kullanma hakkı verirken, başkalarının yeniliği izinsiz kullanmasını engeller. Radikal yenilikler ve çığır açan teknolojik gelişmeler, mevcut patentlerde yer alan teknik bilgilerden türetilmiştir. Otomatik bir sınıflandırma sistemi kullanılarak, ait oldukları teknik sınıfa atanan patentler, araştırmacıların önünü açabilmekte ve yeni buluşlar yaratabilecekleri bir ortam sağlayabilmektedir. Bu çalışma, BERT algoritmasını kullanarak otomatik bir patent sınıflandırma analizi sunmaktadır. Otomatik patent sınıflandırma problemlerinde daha başarılı tahmin doğruluğuna ulaşabilmek için yapılan hiper parametre analizleri bu çalışmada da tercih edilmiştir. Elde edilen sonuçlar literatürdeki sonuçlarla rekabet edecek düzeydedir. Bu çalışmada alt sınıf düzeyinde %58 tahmin doğruluğu elde edilmiştir.

## Patent classification with pre-trained Bert model

### HIGHLIGHTS

- In this study, a solution to the automatic patent classification problem was sought with the Bert model
- The most appropriate parameter combination is tried to be found through hyperparameter analysis
- The result obtained from the study is competitive with the studies in the literature

### Article Info

Research Article

Received: 04.05.2023

Accepted: 05.03.2024

### DOI:

10.17341/gazimmfd.1292543

### Keywords:

Patent classification,  
Bert, deep learning,  
text classification

### ABSTRACT

Patents are documents that help protect innovations in information technologies and grant special rights to the creator of these innovations for a certain period of time. While these rights give the patent owner the right to use the innovation commercially, they prevent others from using the innovation without permission. Radical innovations and ground-breaking technological advances are derived from technical information contained in existing patents. Using an automatic classification system, patents assigned to the technical class to which they belong can pave the way for researchers and provide an environment in which they can create new inventions. This study presents an automatic patent classification analysis using the BERT algorithm. Hyperparameter analyses are also preferred in this study in order to achieve more successful prediction accuracy in automatic patent classification problems. The obtained results were at a level that competed with those in the literature. An accuracy of 58% was achieved at the subclass level.

\*Sorumlu Yazar/Yazarlar / Corresponding Author/Authors : \*yucesoy@gantep.edu.tr, alptekin.durmusoglu@samsun.edu.tr, turkay.dereli@hku.edu.tr / Tel: +90 342 317 2626

## 1. Giriş (Introduction)

Otomatik patent sınıflandırma sistemi, patent metinlerinin ait oldukları teknik alana göre ayrılmasını ve böylece teknik bilgiye daha kolay ulaşılmasını sağlar. Otomatik patent sınıflandırma sistemleri, fikri mülkiyeti tanımlama, kataloglama ve analiz etme sürecini düzene soktukları için önemlidir. Patent başvurularının katlanarak artmasıyla birlikte, manuel sınıflandırma artık mümkün değildir bu yüzden verimli ve güvenilir bir otomatik patent sınıflandırma sistemine çok büyük bir ihtiyaç vardır. Bu sistemler, makine öğrenimi algoritmaları sayesinde çok büyük miktarda patent belgesini kısa sürede inceleyip sınıflandırabilmekte ve gelişen teknolojiler, endüstriler ve ticari eğilimler hakkında yararlı bilgiler sunabilmektedir. Patent metinleri, bilim dünyasındaki genel kabule göre %70 ile %90 arasındaki teknik bilginin yayımlandığı tek teknolojik kaynaktır [1]. Bu yüzden bilimsel makalelere kıyasla, patent meta verilerini kullanarak teknolojik ağlar kurmak ve teknolojinin evrimini incelemek o bilgileri detaylandırmak çok daha kolaydır. Patent metinleri mucitlere, yatırımcılara ve politika yapımcılara değerli bilgiler sağlayarak, onların fikri mülkiyet varlıkları ve pazar fırsatları hakkında bilinçli kararlar vermelerine yardımcı olabilir [2–4]. Genel olarak, otomatik patent sınıflandırma sistemleri, fikri mülkiyet ekosisteminin belkemiğidir ve bilgiye dayalı bir ekonomide ilerlemeyi ve yeniliği destekler.

Teknolojiyi sınıflandırarak bir sistem oluşturmak ve bu yapıyı kullanarak aday patentleri ait oldukları sınıflara atamak, otomatik patent sınıflandırma mantığının en temel özelliğidir. Patentlerin ait olabileceği teknoloji sınıfına uygunluğu, aynı konuda yazılan patentlerin benzerliğine bakılarak belirlenebilmektedir. Araştırmacıların bu süreçte yaşayabilecekleri en önemli sorunlardan biri, önceden belirlenmiş bir teknoloji sınıfının yeni bir buluş ile uyumluluğunun sınırlı olabmesidir. Bazen daha önce hiç rastlanmayan teknolojiye sahip ve alanında çığır açacak bir teknolojik buluş patent alma sürecine girebilmektedir [5]. Bu yenilik nedeniyle, patent araştırmacıları, yeni teknolojilerin önceden belirlenmiş sınırlı sınıflarla uyumluluğunu sağlama konusunda sorun yaşayabilmektedir. Ayrıca teknolojik değişimler nedeniyle yeni buluşlar geçmişte ait oldukları teknoloji sınıflarına ait olmayabilir ve daha uzak sınıflardaki patentlerle benzerlikler gösterebilir. Mevcut sınıflandırma şemaları, sınıf ve alt sınıf düzeylerinde teknolojiye yetişmek için mücadele etmektedir [6–8]. Otomatik patent sınıflandırma sistemlerinin kullandığı bilimsel sınıflandırma araçları güncel literatür ile uyumlu olmalıdır. Bu sayede sınıflandırma sistemlerinin yanlış ve/veya eksik sınıflandırmasından kaynaklanan sorunların çözümüne katkı sağlanabilir.

Patent belgelerinin otomatik olarak sınıflandırılması süreci çok önemlidir. Patentlerin uygun sınıfa atanması işleminde görevli olan patent uzmanları, patentin sadece başlık veya özet kısımlarına bakarak karar vermez. Çoğu durumda istemler ve tarifnameye de bakmak zorundadırlar. Birçok hukuki ve teknolojik terimin bir arada olduğu ve başka bir patentin kapsamına girmemek adına sürekli yeni terminolojiler ile harmanlanarak oluşturulan patent metinlerini manuel olarak anlayıp analiz etmek uzman için hayli zorlayıcı bir süreçtir. Uzmanın bu teknolojileri tam olarak anlayabilmesi için mesaisinin bir kısmını da kendini güncel tutmaya harcaması gerekmektedir. Manuel olarak ilerleyen bu süreç zaman ve efor anlamında büyük bir kaynak israfına yol açmaktadır. Otomatik patent sınıflandırma süreçleri uzmanlara, konu ile ilgilenen patent araştırmacılarına, buluş sahiplerine ve bilim insanlarına araştırmalarını daha verimli yapmaları ve eforlarını başka konulara yönlendirmeleri anlamında yardımcı olmaktadır [9, 10].

Patent dokümanları literatürde bulunan birçok yöntemle analiz edilmektedir. Bunlardan birisi de metin madenciliğidir [11]. Otomatik

patent sınıflandırma yöntemleri, patent meta verilerinin yanı sıra patent metinlerini kullanarak da sınıflandırma işlemini yapabilmektedir. Otomatik patent sınıflandırma sistemleri birçok patent arama adımında kullanılmaktadır. Bu adımlar arasında patent analizi, yeni teknolojilerin araştırılması ve patentlerin uygun sınıflara atanması yer almaktadır. Karmaşık olan bu sistemler teknolojik kategorilerin bulunduğu hiyerarşik bir yapıya sahiptir. Sistem beş seviyeden oluşmaktadır. Bunlar en genelden en özel sınıfa doğru bölüm, sınıf, alt sınıf, grup veya alt grup olarak sıralanabilir [3]. En derin seviyede teknolojinin en spesifik özelliği yer almaktadır. Sistem çok sınıflı ve çok etiketlidir. Bir patent aynı anda birden fazla teknoloji sınıfı altında birden çok etikete sahip olabilir. Ayrıca yoğunluğun bazı kategorilerde yüksek, bazılarında düşük olması da sistemdeki dengesizliğin bir diğer nedenidir. Yapılan araştırmalar dokümanların yaklaşık %80'inin kategorilerin %20'sinde toplandığını ortaya koymuştur [12]. Ayrıca patent metinleri yapısal olarak teknik ve hukuki terimlerin ağırlıkta olduğu terminolojiye sahiptir. Bu durum alanında yetkin olan araştırmacıyı veya patent ofisinde çalışan patent uzmanını, patentleri uygun teknoloji sınıfına atamasını yaparken bile zorlayacak düzeydedir [13].

Patent sınıflandırma alanındaki çalışmalar, buluş sahipleri, patent vekilleri ve araştırmacılar gibi kullanıcılara daha iyi hizmet etmek ve patent sınıflandırma işlemini otomatikleştirmek için yapılmıştır. [14]. Patent sınıflandırma amacıyla oluşturulan ilk patent veri kümesi Wipo-alpha'dır. Bu veri seti, Uluslararası Patent Sınıflandırması (IPC) hiyerarşisine sahip 72.000 alt gruba sahip 46.324 eğitim dokümanı ve 28.926 test dokümanı içeren bir veri setidir [14, 15]. Daha sonra, bu alanda çeşitli veri kümeleri üretilmiştir. Bunlardan en bilineni CLEF-IP (2009-2010-2011) serisidir [16, 17]. Derin öğrenme alanındaki çalışmalar ortaya çıktığında büyük veri kullanımı da eş zamanlı olarak artmıştır. Bu alandaki çalışmalarda kullanılan en önemli parametrelerden biri veri kümeleridir [18]. Önceleri daha az patent dokümanına sahip veri kümeleri ile analizler yapılırken, son yıllarda büyük veri alanında gelişen teknoloji ile birlikte USPTO-2M, USPTO-3M gibi milyonlara dayanan veri kümeleri ile analizler yapılmaya başlanmıştır [19, 20].

Patent sınıflandırma çalışmalarının evrimini incelediğimizde, patentler geleneksel sınıflandırma yöntemleri kullanılarak sınıflandırılmıştır. Çeşitli makalelerde kNN, SVM, naive Bayes, karar ağacı ve lojistik regresyon gibi yöntemler kullanılmıştır [14, 21, 22]. Daha sonra, yapay sinir ağı yöntemleri [23, 24] ve sezgisel yöntemler [25, 26] kullanılarak sınıflandırma çalışmaları gerçekleştirilmiştir. Bilimsel çalışmaların büyük veri üzerinde yoğunlaşmasıyla birlikte CNN [27, 28], RNN ve BiLSTM gibi algoritmaların kullanıldığı çalışmalar literatürde yerini almıştır. Önceden eğitilmiş kelime gömme sistemleri ile birlikte veri kümelerini eğitmek için fazla zaman harcamadan hazır veri kümeleri kullanılarak analizler yapılmış ve bu analizlerde yine ön işleme süreci gerçekleştirilmiştir [29]. Kelime gömme yöntemleri ile vektör uzayları metinleri sayısal verilere dönüştürmektedirler [30]. BERT türevi teknolojilerdeki son gelişmeler sayesinde artık veriler çok fazla ön işleme gereke kalmadan ham haliyle algoritmaya yüklenmekte ve işin büyük bir kısmı önceden eğitilmiş ve çeşitli altyapılarla desteklenmiş algoritmalar tarafından yürütülmektedir [31-33].

Otomatik patent sınıflandırma alanında yapılan çalışmaları birbiriyle kıyaslamak birkaç nedenden dolayı zordur. Bunlardan ilki veri setlerindeki boyutsal farklılıklardır. Bazı çalışmalarda veri sayısı milyonlara ulaşabilmektedir [34]. Bir diğeri ise, problem tanımlarındaki değişkenliklerdir. Patent belgesi yapısı gereği çok etiketlidir. Çok etiketli sınıflandırma yapan çalışmalar olduğu gibi, bazı çalışmalarda bunun aksine ikili veya çok-sınıflı sınıflandırma tercih etmektedir [34–36]. Patent belgelerindeki metin bölümlerinin

seçimi konusunda da farklılıklar göze çarpmaktadır [22]. Sınıflandırmanın ineceği alt kategori seviyesi açısından da farklılıklar gösterebilmektedir. Bazı çalışmalar sadece patent bölümlerini sınıflandırmayı tercih ederken bazı çalışmalarda ise patent sınıflarındaki bölüm, alt-sınıf ya da alt grup düzeyinde sınıflandırma tercih edilebilmektedir [37].

Bu çalışmada, patent belgelerindeki sadece metinsel verileri kullanarak sınıflandırma yapmak amaçlanmıştır. Analizler de önceden eğitilmiş bir BERT modeli kullanılmıştır. Analiz için literatürde sıkça yer alan WIPO-alpha veri kümesi kullanılmıştır.

Bu makalenin geri kalanı aşağıdaki şekilde düzenlenmiştir. İkinci bölümde BERT ve patent sınıflandırmaları hakkındaki yaygın literatür gözden geçirilmektedir. Üçüncü bölümde kullanılan yöntemler açıklanmaktadır. Dördüncü bölümde yapılan çalışma detaylı olarak anlatılmakta ve sonuç bölümünde elde edilen sonuçlar incelenmektedir. Son bölümde ise yapılan çalışmanın bir özeti ve gelecek araştırmalar için öneriler sunulmaktadır.

## 2. Literatür Çalışmaları (Literature Review)

Patentlerin sınıflandırılması, patent belgelerini teknolojik alanlara göre gruplamak için önemli bir uygulamadır. Bu alanda yapılan çalışmalarda farklı veri setleri, metin işleme teknikleri ve klasik ve/veya modern sınıflandırma modelleri kullanılmaktadır. Literatürde, patent belgelerinin başlık, özet, istem ve tarifname gibi metinsel kısımlarının sınıflandırma girdisi olarak kullanıldığı görülmektedir. Bu çalışmaların yanı sıra patent metinlerinin yapısal özelliklerinden yararlanılarak bibliometrik verilerinin kullanıldığı sınıflandırma çalışmaları da mevcuttur. Ayrıca, patent belgelerinin birden fazla etikete sahip olma özelliği dikkate alınarak çok etiketli sınıflandırma çalışmalarının da bulunduğu gözlenmektedir.

Patent sınıflandırması alanında yayınlanan ilk çalışmalardan biri Fall vd. tarafından hazırlanmıştır [38]. Bu çalışma ile birlikte Wipo-alpha veri seti, patent veri araştırmaları dünyasına katıldı. Bu çalışmayı örnek olarak ilerleyen yıllarda birçok farklı çalışma yapılmıştır [25, 39, 40]. Aioli vd. [41] yaptıkları çalışmada tercihi öğrenme modeli ile özet, başlık ve tarifname kombinasyonu (ilk 300 kelime) kullanılarak alt sınıf sınıflandırması yapılmış ve 0,52 doğruluk elde edilmiştir. Wipo-alpha D-bölümü kullanılarak yapılan bir başka çalışmada ise NN yönteminin SVM ve k-NN yöntemleri ile karşılaştırıldığı 46 farklı denemeye yer verilmiştir. Analiz sonuçlarında, özet, başlık ve açıklama bölümlerinin sınıflandırma girdisi olarak kullanılmasının, tüm metinsel verilerin kullanılması arasında anlamlı bir fark yaratmayacağı belirtilmiştir. Wipo-alpha'nın alt kümesinin kullanıldığı çalışmada [42] ana grup düzeyinde sınıflandırma yapılmış ve SVM ve k-NN teknikleri kullanılarak 0,36 doğruluk düzeyi elde edilmiştir.

Makine öğrenmesi kullanılarak metin analizine yönelik yaklaşımlar, son dönemde derin öğrenme çalışmalarının ortaya çıkmasıyla birlikte artış göstermiştir. Yapılandırılmamış metinlerden karmaşık ve anlamlı sonuçlar çıkarabilen bu yöntemler patent sınıflandırmasına da uygulanmıştır. Patentler yarı yapısal özelliklere sahiptir. Bir patentin meta veri bilgilerine (patent numarası, tarih, buluş sahibi, devir alan, ana sınıf ve atıf) ek olarak, çeşitli sınıflandırma çalışmaları analizlerde metinsel bilgileri (başlık, özet, istem ve tarifname) de kullanmıştır [43-45]. Örneğin Hain vd. [6] yaptıkları çalışmada, patent sınıflandırma yöntemlerini dört farklı başlık altında toplamışlardır. İlk olarak patentin metinsel bilgileri anahtar kelime tabanlı çalışmalarda (kelime frekansları ve eş-oluşum oranları incelenerek sınıflandırma) analiz edilmiştir. Kullanılan ikinci yaklaşım ise; özne-eylem-nesne (SAO) (dil bilimsel olarak işlenmiş metinlerde özne ve nesne

arasındaki ilişkiye dayalı olarak patentlerin sınıflandırılması) yaklaşımlarıdır. Üçüncü yaklaşım belirli bir alandaki kavramları ve bunların ilişkili olduğu ilişkileri tanımlayan ontolojik sınıflandırma yöntemlerinin kullanılmasıdır. Dördüncü olarak, makine öğrenmesi ve derin öğrenme tabanlı yaklaşımlardır (büyük veri kümeleri ve dönüştürücülerini analize dahil ederek sınıflandırma). Tüm bu yaklaşımlar son dönemde literatürde en sık karşılaşılan yöntemler arasında sıralanabilir.

Geleneksel metin analizinde sıkça kullanılan tfidf gibi yöntemler, belgelerdeki kelimelerin terim frekans vektörünü esas alır. Ancak bu yöntemler, kelimenin anlamını veya benzer belgelerdeki kelimelerle olan ilişkisini göz önünde bulundurmaz. Bu sorunu gidermek için, patent sınıflandırma problemlerinde çeşitli gömme yöntemleri kullanılmıştır. DeepPatent isimli çalışma, derin öğrenme alanında en çok alıntı yapılan çalışmalardan biridir. Çalışmada kurulan algoritma ESA (Convolutional Neural Network, Evrişimli Sinir Ağı) yöntemi ve word vektör gömme tabanlı (word2vec) vektörleri birleştirmektedir. DeepPatent çalışmalarında yazarlar alt sınıf düzeyinde sınıflandırma yapmak için Clef-IP (Conference and Labs of the Evaluation Forum) patent verilerini (2.000.147 kayıtlı USPTO-2M) kullanmışlar. Bu iki veri seti ile yapılan analizlerde sonuçlar diğer algoritmalarından (Decision Tree, Naive Bayes, Random Forest) üstün çıkmıştır. Algoritmanın F1 skoru %73,88 kesinlik ile %43 olmuştur [12]. Bu çalışmada klasik metin sınıflandırma algoritmalarının yanı sıra kelime gömme ve ESA yöntemleri kullanılmış ve patent metinlerinin farklı bölümleri üzerinde analizler yapılmıştır. Başlık, istem ve tarifname kısımlarının farklı kombinasyonlarının girdi olarak kullanıldığı çalışmada patent belgelerinin başlık ve özet bölümlerinden alınan ilk 100 kelimenin iyi performans sağlayabileceği sonucuna varılmıştır. Gömme vektörleri ile yapılan analizler, patent belgelerinde istenen yüksek doğrulukta sınıflandırma performansına ulaşamadı. Çünkü patent belgeleri farklı karmaşık bir jargona sahiptir. Risch ve Krestel [46] patent alanındaki kelimeler için alan-özüğü gömme elde etmek için 5 milyondan fazla patent belgesi üzerinde fastText yaklaşımını kullanarak bir kelime gömme vektörünü eğittiler. Çift yönlü Gru mimarisi ile bu vektörü besleyerek sınıflandırma yaptılar.

BERT'in çift yönlü bağlamsal bilgiyi yakalama, mimari değişikliğe gerek duymadan ince ayar yapabileme ve patent belgelerini doğru anlama ve sınıflandırma gibi avantajları vardır [32, 47]. Otomatik patent sınıflandırma çalışmalarında dil anlama ve öğrenmede üstün olan BERT algoritmaları yaygındır. Bu çalışmalardan bazıları kelime gömme ve Bert algoritmalarının çeşitli varyasyonlarını birleştirmişlerdir. Örneğin Roudsari vd. yaptıkları çalışmada [48], USPTO-2M ve M-patent veri setlerini özet ve başlık bölümleri girdi verisi olarak almışlardır ve IPC alt sınıfı düzeyinde ön eğitimli BERT, XLNet, RoBERTa ve ELECTRA algoritmaları kullanarak çok etiketli bir patent sınıflandırma işlemi gerçekleştirmişlerdir. Analiz sonuçları, CNN, LSTM, BiLSTM ve CNN-BiLSTM derin öğrenme modelleri ile karşılaştırılmış ve önceden eğitilmiş dil modellerinin daha iyi performans gösterdiğini ve fastText gömmenin word2vec gömmeye göre daha iyi olduğunu ortaya çıkarmıştır. XLNet algoritması, M-patent veri setinde 72,08 F1 değeri ve USPTO-2M veri setinde 63,33 F1 değeri ile önceden eğitilmiş diğer modellerden daha iyi performans göstermiştir. Mevcut çalışma PatentNet çalışmasından [48], kullandığı veri seti çeşidi açısından (USPTO-2M ve M-patent veri setleri), yapılan sınıflandırmanın çeşidi açısından (çok etiketli patent sınıflandırma) ve kullandığı yöntemlerden dolayı (BERT, XLNET ve ELECTRA) farklılık göstermektedir. CPC (Cooperative Patent Classification) veri setine Sentence-Bert (SBERT) algoritmasını uyguladığı bir başka çalışmada, 1.492.294 patentte alt sınıf düzeyinde %54 doğruluk ve %66 F1 puanı elde edilmiştir. 10000 cümlelerin SBERT algoritması ile analizi 5 sn, aynı işlem BERT ve RoBERTa ile yapıldığında 65 saat sürmüştür [20]. Lee vd. yaptıkları çalışmada yeni ve önceden eğitilmiş BERT tabanlı transformör

modeli PatentBERT adıyla patent sınıflandırması için geliştirilmiştir. USPTO-3M (3.050.615 patent) veri setinin CPC alt sınıfı seviyesindeki patentler için kullanılmıştır. Girdi verisi olarak patent başlığı, özeti ve istemleri kullanılmıştır ve %44,75 F1 değerine ulaşılmıştır [19]. Wang vd. yaptıkları patent haritalama çalışmasında ise Bert ve Triz yöntemleri kullanılmıştır. Derwent veri tabanından patent verileri toplanmıştır. Benzer teknoloji öğelerini aynı başlıklarda toplayarak bir vektör uzayı kurulmuştur. Triz var olan teknolojileri bulmaya yardımcı olurken Bert sınıflandırma aşamasında devreye girmiştir[49]. Zhipeng ve Zheng [50] yeni teknolojilerin tasarlanmasında bazı soruları çözmek için bir derin öğrenme modelinin çıktı öznelik vektörleri alıntılarının fiziksel anlamı ile birleştirdiler. Patent metinlerine yapılan atıfları vektör olarak kabul ederek yaptıkları analize katmışlardır. Bert modelini analizde kullanan çalışmada patentlerin atıf ilişkileri ve vektörleri ilk vektörlerle büyük benzerliklere sahip olan patentlerin ileriye dönük alıntılarını yoluyla mucitlerin yeni teknolojiler tasarlamasına yardımcı olacak bir yaklaşım geliştirilmiştir. Choi ve ark yaptıkları çalışmada[51], derin bir patent haritalama modeli önerdiler. Bu model 2 bölümden oluşmaktadır. Birincisi bir transformatör- kodlayıcıdır (transformer encoder) [52], ve ikincisi, grafik gömme (graph-embedding) işlemi olan Diff2Vec'dir. Metinsel bilgiler ve atıf-grafik bilgi verilerinden yararlanılarak derin patent haritalama modeli ile sınıflandırma sorununa çözüm bulunmaya çalışılmıştır. Çalışmanın patent veri kümesi Google BigQuery'den elde edilmiştir. Analiz sonuçları APL ve PatentBERT ile karşılaştırılmıştır[51].

Bu çalışma, otomatik patent sınıflandırma alanında sıkça karşılaşılan DeepPatent ve PatentBert çalışmalarından farklı yönleri olan bir çalışmadır. DeepPatent makalesinde USPTO 2M ve CLEF-IP veri setlerini kullanarak ESA ve kelime gömme tabanlı bir algoritma geliştirilmiştir. Çalışmada patent metinlerinin sadece başlık ve özet kısımları kullanılmış ve %73,88 doğruluk ve %55,09 F1 skoru elde edilmiştir. PatentBert makalesinde ise, USPTO 3M veri setinin yanı sıra aynı zamanda DeepPatent'in kullandığı USPTO-2M veri seti de kullanılmış analizler karşılaştırmalı olarak incelenmiştir. Patent metinlerinin sadece istem kısımlarının yeterli olduğu savunulmuştur. Her iki çalışmada, metin verilerinin yanı sıra patent etiketlerini de analize katmaktadırlar. Wipo-alpha veri setini kullanan diğer çalışmalara bakıldığında, Gomez vd. çalışmasında Wipo-alpha ve Wipo-de patent veri setleri kullanılmıştır. Bu çalışmada, patentin hem etiketi hem de metin bölümünü girdi olarak alınmıştır. NB, SVM, LR ve KNN yöntemleri kullanılmış ve Wipo-alpha alt sınıf seviyesinde ulaşılan en iyi sonuç (doğruluk=0,533), lojistik regresyon ve tarifname girdisi ile elde edilmiştir. Abdelgawad vd. çalışmasında ise, farklı kelime gömme yöntemleri (glove, fasttext) ESA metodu ile birlikte kullanılmıştır. Alt sınıf seviyesinde %55,02 doğruluğa ulaşılmıştır. PatentBert ve DeepPatent çalışmalarıyla birçok açıdan farklılık göstermektedir. Bunlar 1) patent sınıflandırması için Wipo-alpha veri setinin kullanılması 2) patent belgesinin sadece metin kısmı kullanılarak sınıflandırma performansı elde edilmeye çalışılması 3) Bert modelinin hiper parametreleri analiz edilerek sınıflandırma performansı üzerindeki etkilerinin tartışılması olarak sıralanabilir.

Bu çalışmanın literatüre katkıları şöyle sıralanabilir. İlk olarak, literatürde yaygın olarak kullanılan ve patentin sadece özet ya da istem bölümünün girdi verisi olarak kullanılmasının yeterli olduğu düşüncesi bu çalışmanın ulaştığı sınıflandırma performansı sonucuna göre geçerliliğini yitirmiştir. İkinci olarak birçok hiperparametre analizi ile gerçekleştirilen denemelerden elde edilen sonuçlar, bu alanda analiz yapmak isteyen araştırmacılar için bir başlangıç noktası oluşturarak bilgisayar zamanı ve Gpu gibi verilerin gereksiz kullanımını engelleyerek zamandan tasarruf ettirecektir. Üçüncü olarak, gerçekleştirilen alt sınıf düzeyindeki sınıflandırma analizlerinin elde ettiği doğruluk değeri literatürdeki benzer çalışmaların ulaştığı doğruluk değerini geçmiştir.

### 3. Materyal ve Yöntem (Material and Method)

#### 3.1. BERT (Bidirectional Encoder Representations from Transformers) (Transformatörlerden Çift Yönlü Kodlayıcı Gösterimleri)

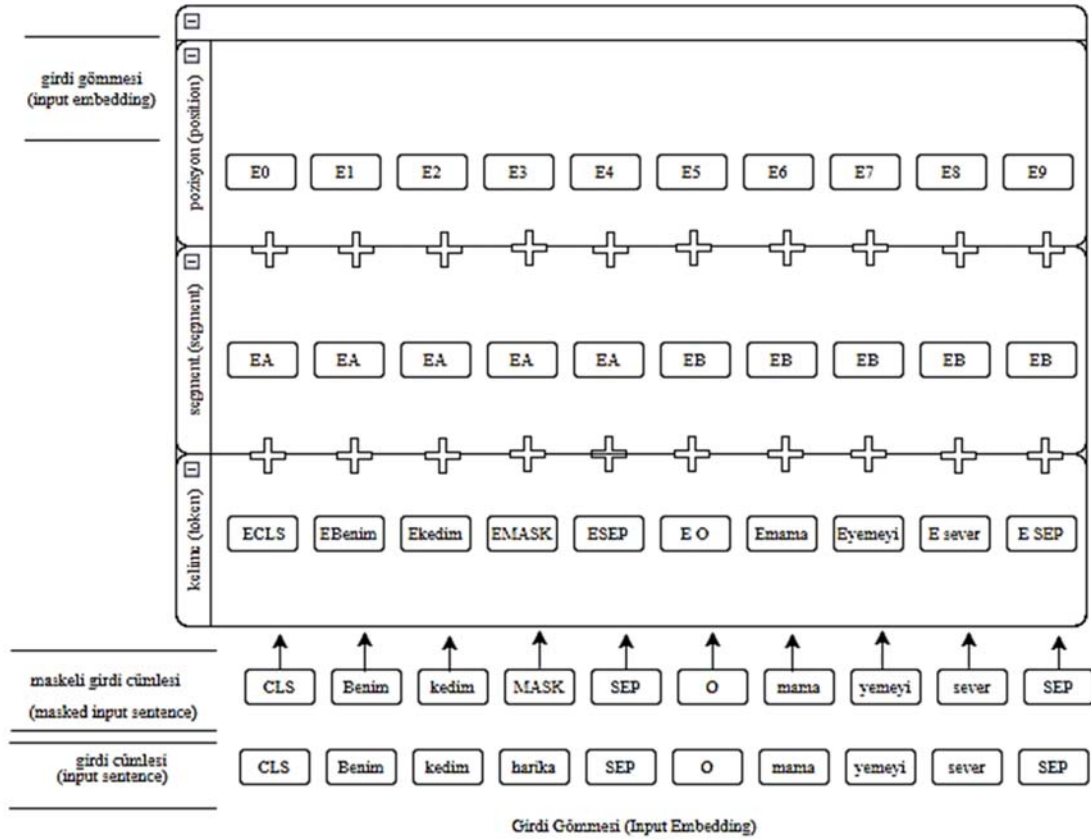
BERT (Bidirectional Encoder Representations from Transformers (Transformatörlerden Çift Yönlü Kodlayıcı Gösterimleri)), Google araştırmacıları tarafından 2018'de önerilen bir dil temsil modelidir. Bert, daha önce etiketlenmemiş metinlerin tüm katmanlarını analiz eder. Hem sağ hem de sol bağlamları kullanarak çift yönlü çalışır ve yalnızca bir çıktı katmanı kullanan önceden eğitilmiş Bert modeli güncel sorunlara çözüm üretebilir. Bert modelini uygulamak için, her model için algoritmanın mimarisini değiştirmek gerekli değildir [32]. Doğal dil işleme problemi ile ilgili 11 farklı güncel alan ve konuda test edilen Bert modeli, bu alanların birçoğunda tahmin puanlarını bir üst seviyeye çıkararak kendini kanıtlamayı başarmıştır [32, 53]. Bert, önceden etiketlenmemiş büyük miktarda ham veri, maskeli sözcüklerin tahmini ve sonraki cümleyi bulma kullanılarak eğitilmektedir. Daha sonra etiketli veri kümesinde ince ayar yapılabilmektedir. Önceden eğitilmiş Bert modeli geliştirme süreci, gereken büyük miktarda veri nedeniyle zordur. Bert modelinin geliştiricileri, eğitim için 110 milyar parametrelilik 12 katman kullanmışlardır. Yoğun yapıları geliştirmek için büyük miktarda veri gerekmektedir [54]. Bert'in model mimarisi Vasmani vd. [52], tarafından yayınlanan transformatör yapısından birkaç farkla uyarlanmıştır. Bu yapının iki tarafı vardır, solda kodlayıcı ve sağda kod çözücü bulunur. Kodlayıcı giriş dizisinin temsillerini oluşturur ve kod çözücü bu temsilleri çıkış dizilerine eşler. Kodlayıcı ve kod çözücü, her katmanın alt katmanlara sahip olduğu birkaç katmana sahiptir. Kodlayıcının katmanları 2 alt katmana sahipken, kod çözücünün 3 alt katmanı vardır. Yapıda üç farklı dikkat türü bulunmaktadır. Bunlar Çoklu kafalı dikkat (multi-head attention), Maskeli Çok Kafalı (Başlı) Dikkat (masked-multihead attention) (sadece kod çözücüde) ve ileri beslemeli ağlardır (feed forward network). Çoklu kafalı dikkat yapısında anahtarlar, öz dikkatten gelen değer sorguları ve bir önceki katmanın çıktısı, paralel olarak dikkat işlemini gerçekleştirmek için doğrusal olarak gönderilir. Maskeli Çok Kafalı dikkat sadece kod çözücüde bulunan ilk alt katmandır. Görevi, henüz bilinmemesi gereken belirli bir konumdaki tahminleri önmektir. İleri Besleme Ağı, her pozisyonda iki doğrusal dönüşümü ayrı ayrı ve eşit olarak uygulamaktan sorumludur. Dikkat işleme ve bilgi giriş katmanlarından önce, konumsal kodlama bir dizideki bir tokenin konumu hakkında göreceli bilgi sunar. Transformatör bu bilgiye dayanarak bir token sırası kullanır. Katmanlarda dikkat ve dönüşüm işlemi gerçekleştirildikten sonra bir diğer doğrusal dönüşümü gerçekleştiren Lineer ve çıktıyı olasılıklara döndüren Softmax dönüşümleri gerçekleştirilir.

Bert, iki yönlü eğitim kullanarak inovasyon gerçekleştirir. İki ön eğitim görevi vardır: maskeli dil modelleme (masked language modeling(MLM)) ve sıradaki cümle tahmini (next-sentence prediction (NSP)). Bert algoritması eğitim aşamasında MLM tekniğini kullanmaktadır. Bu teknikte maskelenen kelimeler vardır. Bu kelimeler daha önce maskelenmemiş kelimeler tarafından tahmin edilirler. Bu işlemin temel amacı cümle içindeki kelimeleri incelemek ve tahmin işlemi sağlamaktır. MLM, her dizedeki kelimelerin rastgele seçilen bir yüzdesini (%15) gizler ve geri kalanını sisteme gönderir. Ancak, bu %15 token'ın yalnızca %80'i [MASK] token'ı ile değiştirilmiş, %10'u rastgele başka bir token ile değiştirilmiş ve %10'u değişmeden kalmıştır. Sonraki Cümle Tahmini (NSP) de cümleler arasındaki modeli eğitir ve her parça birkaç cümleden oluşan cümle çiftleri arasındaki ilişkiyi belirler. Bir cümlenin öncesinde ve sonrasında bulunan cümleler ile olan yapısal ve anlamsal ilişkileri ortaya konulur.[55] Bert NSP işleminde girdileri ikiye böler. Eğitim

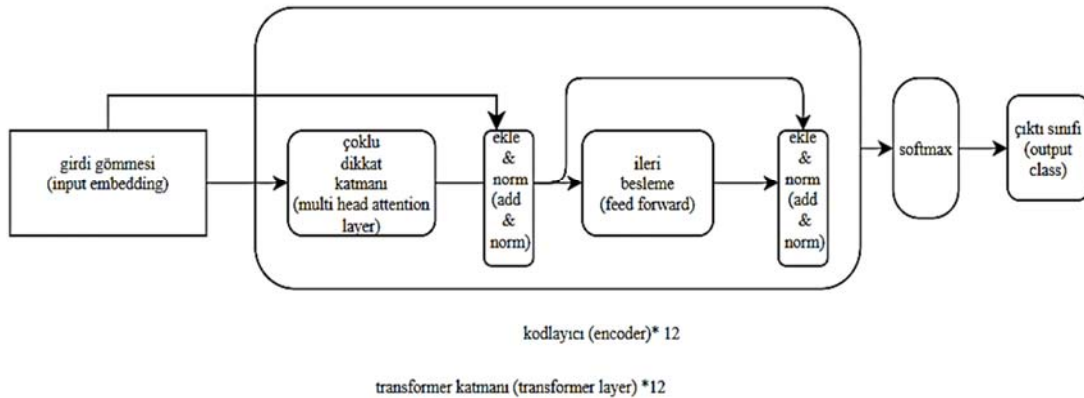
aşamasında ikinci cümledeki birinci cümleyi takip ettiği ve bağlı olduğu öğretilirken, test aşamasında ikinci cümle rasgele seçilerek ilk cümleye bağlı olup olmadığını tahmin etmesi istenir[56]. BERT, Wu ve diğerleri [57] tarafından geliştirilen WordPiece gömülerini kullanır. BERT ilk olarak cümleleri ayırmak için ([SEP]) kelimesini kullanarak cümleleri ayırır. İkinci olarak, her bir tokenin ait olduğu ifadeyi tanımlamak için [CLS] kelimesini kullanır [32]. Şekil 1’de CLS kelimesi sınıflandırma işlemi için kullanılan tokendir. Ayrıca basit sıralı iki cümleyi ayırmak için SEP tokeni kullanılmaktadır. 3 farklı girdi gömmesi Bert’in kullandığı gömme çeşitleridir. Kelimeler token gömmesi, kelimelerin cümle içindeki yeri için pozisyon gömmesi ve tokenin hangi cümleye ait olduğunu anlamak için de segment gömmesi kullanılır. Örneğin “kedim” kelimesi “E<sub>kedim</sub>” token

vektörü ile token gömmesinde yer alır ve onun cümle içinde hangi pozisyonda yer aldığı E<sub>3</sub> ile belirtilmektedir. Hangi cümleye ait olduğunu bilmek için de segment gömmesinde E<sub>A</sub> vektörü ile ifade edilmektedir. Bert belli oranda kelimeyi maskeler burada “harika” kelimesi “mask” kelimesi ile maskelenmiştir. Bert’in çıktısı bu kelimeyi bulmaya yönelik tahmin işlemi olacağı için Çıktı vektörü de harika kelimesi dördüncü sırada olduğu için Q<sub>4</sub>’tür. Daha sonra ileri beslemeli sinir ağları ile ve softmax aktivasyon fonksiyonu ile kelimelerin olasılık değerlerini bulur. Burada Bert farklı olasılıklarda “harika”, “güzel”, “şirin” gibi kelimeleri önerebilir.

Şekil 2’de BERT mimarisinin ikinci aşaması olan transformer katmanı gösterilmektedir. Bu aşamada, girdi gömme katmanından



Şekil 1. Bert’in girdi temsili (Input representation of Bert)



Şekil 2. Bert’in transformer aşaması (Transformer phase of Bert)

gelen veriler işlenerek çıktı verisi üretilir. Bu işlem 12 defa tekrarlanır. Transformer, çok kafalı dikkat mekanizması sayesinde gelen veriyi hem sağdan sola hem de soldan sağa doğru okur ve aynı anda birden fazla hesaplama yapar. Bu şekilde model eğitilir. Burada, dikkat mekanizması, işlenen her veriyi saklı katmanlara ve kodlayıcıya aktararak veri kaybını önler[52, 58].

Bert, iki farklı veri setinin (BookCorpus (800 milyon kelime) ve İngilizce Wikipedia (2500 milyon kelime)) kombinasyonu kullanılarak toplam 16 GB ile eğitildi. BERT'in performansı birçok araştırmacıyı bu alanda çalışmak konusunda heveslendirdi ve onlarda benzer mimaride modeller oluşturdular. Böylece, yeni bir Bert ailesi ortaya çıktı bunlardan bazıları CamemBERT, RoBERTa, RuBERT, DistilBERT ve AIBERT'dir. Bert ailesindeki elemanların birbirleriyle ve BERT ile bazı farklılıkları bulunmaktadır. Örneğin, RoBERTa geliştiricileri, modelin eğitim süresini, veri kümesi boyutunu ve parti sayısını artırmanın performans üzerinde olumlu bir etkisi olduğunu ve bir sonraki cümle tahmini amacının kaldırılarak dinamik maskeleme kullanılması gerektiğini savunmuşlardır [33]. RoBERTa, BERT'ten on kat daha büyük bir veri kümesi kombinasyonu ile eğitilmiştir [53]. DistilBERT, Bert modelinin %40 daha küçük ancak %60 daha hızlı bir versiyonudur[30]. Bert'in becerisinin %97'sini sağlamaktadır [31]. AIBERT, BERT ve DistilBERT ile aynı veri kümelerini kullanır. RoBERTa'nın temel amacı performansı artırmak iken, DistilBERT daha çok hız odaklanmaktadır. Buna karşın, AIBERT bu ikisini aynı anda başarmayı hedefler. Sonraki cümle tahmini yerine cümle sırası tahmini yapar [59]. CamemBERT, RoBERTa ile aynı yapıya sahiptir ve Fransızca için geliştirilmiştir [60]. BERT 12 katman, 768 gizli katman ve 12 öz dikkat başlığı kullanır. Bert-base 110 milyon parametreye sahipken[61], BERT-large 340 milyon parametreye sahiptir (24 katman 1024 gizli katman 16 öz dikkat kafası). RoBERTa-base ve RoBERTa large'in katman yapıları BERT ile aynı olsa da parametre sayıları farklıdır, sırasıyla 125 milyon ve 355 milyon parametreye sahiptirler. [33]. DistilBERT-base yapısındadır ve 66 milyon parametreye, altı katmana, 768 gizli katmana ve 12 öz dikkat yapısına sahiptir. Buna karşılık, AIBERT'in temel modelinde 12 milyon parametre bulunurken, büyük modelde 18 milyon parametre vardı. Katman yapısı RoBERTa gibidir [59].

Bu çalışmada BERT'in çok sınıflı patent sınıflandırmasına yönelik performansı incelenmiştir. Farklı hiper parametreler kullanılarak tahmin doğruluğu artırılmaya çalışılmıştır. Ayrıca patent belgesinin farklı metinsel içerikleri ve onların kombinasyonları girdi verisi olarak kullanılmıştır. Bu yapılan çalışmalar ile Bert algoritmasının farklı uzunluktaki metinlerde göstereceği performans gözlenmiştir.

### 3.2. Çok Sınıflı Patent Sınıflandırma (Multi-Class Patent Classification)

Patent belgeleri hem çok sınıflı hem de çok etiketli olma özelliğine sahiptir. Çok etiketli olmasının temel nedeni bir patentin aynı anda birden çok teknolojiyi bir arada bulundurabilmesidir. Çok sınıflı olması da uluslararası patent sınıflandırma sisteminin temel yapısının

birden fazla alt sınıfa ayrılmış (IPC (Uluslararası patent sınıflandırması) de 76422 tane alt grup bulunmaktadır.) olmasından kaynaklanır [62]. IPC hiyerarşisi 8 bölümden oluşur[63]. Örneğin "sabit yapılar" (Bölüm E) veya "makine mühendisliği; aydınlatma; ısıtma; silahlar; cephaner" (Bölüm F). Tablo 1, IPC'nin sekiz bölümünün tümüne genel bir bakış sunmaktadır. Her bölüm, birden çok sınıfa (örn. E01, E02, E03) dahil edilmiştir ve her sınıf, E01B gibi çok sayıda alt sınıfa bölünmüştür. Her alt sınıfın birden fazla ana grubu vardır (örn. E01B 7/00, E01B 29/00) ve çoğu ana grup için E01B 29/00, E01B 29/02 gibi ek alt gruplar vardır. Bu hiyerarşinin örnek bir tanımlaması Tablo 2'de listelenmiştir.

**Tablo 1.** IPC' nin sekiz bölümünün tümüne genel bakış (Overview of all eight sections of the IPC)

| IPC Bölümleri | Patent Bölümlerinin Tanımı                                  |
|---------------|---|
| A             | İnsani ihtiyaçlar   |
| B             | İşlemlerin uygulanması; taşıma                              |
| C             | Kimya; metalurji  |
| D             | Tekstil; kâğıt  |
| E             | Sabit yapılar   |
| F             | Makine mühendisliği; aydınlatma; ısıtma; silahlar; mühimmat |
| G             | Fizik   |
| H             | Elektrik  |

Otomatik patent sınıflandırmasının en önemli sorunu, patentlerin yanlış sınıflandırılmasıdır. Bu yanlışlık, patentlerin mahkemelerde hükümsüz kılınması veya patentlerin uygulanmasının engellenmesi gibi çeşitli sorunlar yaratabilmektedir. Bu tür sorunların yaşanmaması için otomatik patent sınıflandırma sistemlerindeki hataların en aza indirilmesi gerekmektedir.

Bu çalışma, çok sınıflı patent sınıflandırma probleminin çözümüne yönelik bir araştırmayı içermektedir. Patent belgelerinin uygun teknoloji alt sınıflarına atanmasında yapılan tahmin doğruluğunun iyileştirilmesi amaçlanmaktadır. Önceden eğitilmiş ve cümleleri hem sağdan sola hem de soldan sağa olacak şekilde çift yönlü okuyan bir dil modeli kullanmanın sınıflandırma performansına etkisini gözlemlenmiştir. Bunun yanında, parametreler hiper parametrik analizler yapılmasının tahmin doğruluğuna etkisi incelenmiştir. Bu parametrelerden

öğrenme oranı, hiper parametre analizlerinde en sık kullanılan araçlardan biridir. Modelin performansına doğrudan etki eder. Kayıp fonksiyonlarından aldığı geri dönüşlere göre adım genişliklerini ayarlar. Bu adım genişliğinin büyük tutulması hedeften uzaklaşmaya sebep olurken, adım genişliğinin dar tutulması ulaşılmak istenen noktaya daha yavaş ulaşılmasına ve zaman maliyetinde de artış sonucuna neden olacaktır[64]. Döngü büyüklüğü eğitim verileri üzerinden bir tam geçişe karşılık gelir. Küme büyüklüğü ise bir GPU tarafından bir eğitim adımında kullanılan eğitim örneklerinin sayısıdır [65].

**Tablo 2.** IPC sıra kodu örneği (IPC sequence code example)

| IPC Sıra Ağaç Kodu  | Tanımlama   |
|---------------------|---|
| Bölüm E             | Sabit yapılar   |
| Sınıf E01           | Yol, demiryolu veya köprü inşaatı   |
| Alt sınıf E01B      | Kalıcı yol; kalıcı yol araçları, her türlü demiryolu yapmak için makineler  |
| Ana grup E01B 29/00 | Rayların döşenmesi, yeniden inşası veya alınması; Bunun için aletler veya makineler   |
| Alt grup E01B 29/02 | Birleştirilmiş rayların, birleştirilmiş anahtarların veya birleştirilmiş geçişlerin taşınması, döşenmesi, kaldırılması veya yenilenmesi |

### 3.2.1 Ön- işleme (Pre-processing)

Analizde kullanılacak patent verilerinin girdi metnini önceden eğitilmiş dil modelinin girdi formatıyla eşleşecek şekilde hazırlamak için basit bir ön işleme adımı gereklidir. Bu çalışmada standart metin temizleme prosedürleri kullanılmıştır. Bu prosedürlerden bazıları, tüm metni alçaltma, noktalama işaretlerini kaldırma, HTML bağlantılarını/e-posta adreslerini kaldırma ve sayıları kaldırma gibi adımlardır [47, 66]. Bert algoritmasının genel metin ön işleme adımlarında ilk sırada veri tokenizasyonu bulunur. Bu aşamada, ham metinde bulunan tüm kelime/kelime grupları alt kelimelere ve noktalama işaretlerine ayrılır. Bu adım, tanımlanabilecek daha küçük bir girdi kümesi sağlar ve önceden eğitilmiş modelin sözlük altyapısına uygundur. İkinci adım, sınıflandırma tokenlerini [CLS] ve ayırıcı tokenleri [SEP] eklemektir. Tokenler, önceden eğitilmiş tablodaki indisler kullanılarak sayısal değerlere dönüştürülür. Bir patent metninden alınan tokenizasyon örneği Şekil 3'de gösterilmektedir.

Bu segment, hedef cümleyi bulmak için etrafındaki önceki ve sonraki cümleleri de kontrol eder. Yani bir cümle içerisine kelime gömme ile oluşturulan skip-gram modellerinin bir uyarlamasıdır. Üçüncü tür ise konum gömmedir. Gömme, bir kelimenin cümle içindeki konumunu hafızaya alır. Böylece göreceli konum bilgisi saklanmış olur [67]. Bu durum LSTM'nin uygulama mantığına benzer.

### 3.2.2 Veriseti (Dataset)

Bu çalışmada Wipo-alpha patent veri seti kullanılmıştır [15]. Bu veri seti 75.250 patent verisinden oluşmaktadır. IPC kategorilerinin ilk dört seviyesini içermektedir. 8 bölüm, 114 sınıf ve 451 alt sınıf ve 4.427 ana gruptan oluşmaktadır. XML formatında hazırlanmıştır. Şekil 4'te Wipo-alpha veri setindeki bölümlerin patent sayılarına ilişkin dağılım grafiği bulunmaktadır.

### 3.2.3 Değerlendirme Ölçütleri (Evaluation Metrics)

Bu çalışmadaki modellerinin performansı Eş.1-4 arasındaki formüllere göre belirlenmektedir. Sonuçların yorumlanması için

yapılan değerlendirme işleminde doğruluk, F1, duyarlılık ve kesinlik değerleri kullanılmaktadır. Doğruluk değeri tüm sonuçların yer aldığı karmaşık matrisinden genel bir sonucu gösteren doğru olarak tahmin edilen patent sayısının tüm tahmin sayısına bölünmesi ile bulunmaktadır. Kesinlik değeri, doğru olarak tahmin edilenlerin patent sınıflarının gerçekte de doğru olup olmadığını belirtmektedir. Duyarlılık değeri, doğru sınıfın ne oranda doğru tahmin edildiğini ifade etmektedir. F1 değeri, duyarlılık ve kesinlik değerlerinin harmonik ortalamasının alınması ile bulunan sonuçtur. Bu çalışmanın değerlendirme ölçütleri, kesinlik (precision) duyarlılık (recall), F1 ve doğruluktur (accuracy) [68].

$$\text{Doğruluk} = \frac{DP+DN}{DP+DN+YP+YN} \quad (1)$$

$$\text{Kesinlik} = \frac{DP}{DP+YP} \quad (2)$$

$$\text{Duyarlılık} = \frac{DP}{DP+YN} \quad (3)$$

$$F1 = 2 \times \frac{\text{kesinlik} \times \text{duyarlılık}}{\text{kesinlik} + \text{duyarlılık}} \quad (4)$$

Doğru Pozitif (DP) ve Doğru Negatif (DN) modelin doğru tahmin ettiği alanlardır, Yanlış Pozitif (YP) ve Yanlış Negatif (YN) modelin yanlış tahmin ettiği alanlardır

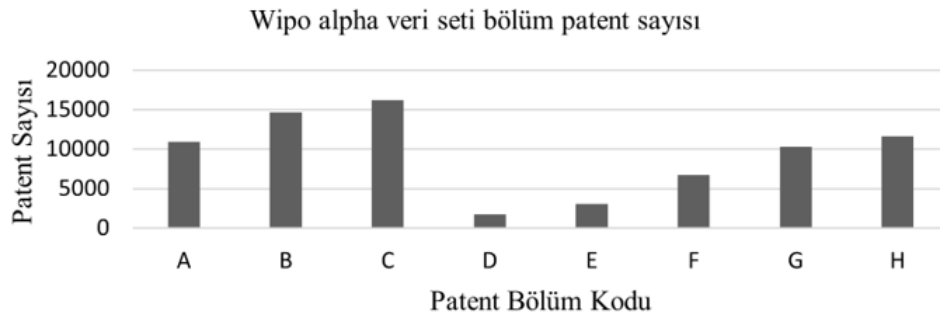
## 4. Sonuçlar ve Tartışmalar (Results and Discussions)

Bu bölümde deneysel sonuçlar sunulmaktadır. Wipo alpha patent dokümanları üzerinde alt sınıf seviyesinde bir dizi patent sınıflandırma deneyi gerçekleştirilmiştir. Bu deneyler şu şekilde özetlenebilir: sınıflandırma performansının farklı öğrenme oranlarındaki değişimi, farklı yığın büyüklüklerindeki değişimi ve patent metinlerinin ayrı ayrı girdi değeri olarak kullanılmasının sınıflandırma performansına etkisi incelenmiştir. Çalışmada kullanılan hiper parametre değerleri Tablo 3'de özetlenmiştir. Tablolarda bulunmayan ancak bu çalışmada yapılan tüm deneylerde

Tokenize the first sentence:

```
[[CLS], 'an', 'apparatus', 'for', 'pre', '##fa', '##bri', '##cating', 'furniture', 'having', 'shelves', '(', '2', ')', 'or', 'guide', 'rails', '(', '60', ')', 'for', 'drawers', 'is', 'disclosed', ',', 'the', 'apparatus', 'comprises', ',', 'two', 'side', 'plates', '(', '10', ')', 'having', 'one', 'or', 'two', 'more', 'flutes', '(', '14', ')', 'formed', 'vertically', 'in', 'the', 'inside', 'surface', '(', '12', ')', 'thereof', ',', 'guide', 'bars', '(', '20', ')', 'being', 'inserted', 'in', 'the', 'flutes', '(', '14', ')', 'and', 'having', 'a', 'vertical', 'guide', 'groove', '(', '22', ')', 'with', 'a', 'slit', '(', '28', ')', ',', 'and', 'supporting', 'members', '(', '40', ')', 'being', 'installed', 'in', 'the', 'guide', 'bars', '(', '20', ')', 'across', 'the', 'slit', '(', '28', ')', 'and', 'supporting', 'the', 'shelves', '(', '2', ')', 'or', 'said', 'guide', 'rails', '(', '60', ')', 'for', 'drawers', 'there', '##on', ',', 'the', 'supporting', 'members', '(', '40', ')', 'can', 'be', 'moved', 'up', 'and', 'down', 'along', 'the', 'guide', 'bar', '(', '20', ')', 'and', 'be', 'fixed', 'to', 'a', 'desirable', 'position', 'thereof', ',', '[SEP]]
```

Şekil 3. İlk cümlelerin tokenlerine ayrılması (Splitting the first sentence into tokens)



Şekil 4. Wipo-alpha veri seti bölüm patent sayısı dağılımı (Distribution of Wipo-alpha dataset section parts)



hiç değiştirilmeden kullanılan başka bazı hiper parametreler vardır. Örneğin tüm deneylerde optimize etme fonksiyonu olarak Adam optimizasyonu kullanılmıştır. Maksimum dizi uzunluğu 128'dir. Aşırı öğrenmeyi engellemek için kullanılan atlama oranı (dropout rate) 0,1 olarak alınmıştır.

**Tablo 3.** Çalışmada kullanılan farklı hiper parametre değerleri (Different hyper parameter values used in the study)

| Hiper parametre               | Değer   |
|-------------------------------|---|
| Öğrenme oranı (Learning Rate) | 1e-5, 2e-5, 3e-5, 4e-5, 5e-5  |
| Girdi Türü (Input Type)       | Başlık, Özet, İstem, Tarifname, Tam metin ve çeşitli kombinasyonlar |
| Yığın (küme) (Batch Size)     | 16, 32, 64, 128   |

Bu çalışmanın Google Colaboratory ile -51 GB RAM'e sahip NVIDIA Tesla V100 GPU üzerinde gerçekleştirilmiş ve veriler python programlama dili kullanılarak analiz edilmiştir.

#### 4.1. Farklı Yığın Büyüklüğüne Göre Sınıflandırma Performansındaki Değişiklikler (The Changes in Classification Performance According to Different Batch Sizes)

Yığın büyüklüğü, derin öğrenme sistemlerinde ilgililenen önemli hiper parametrelerden biri olarak kabul edilmektedir. Bu çalışmada, yığın büyüklüğündeki değişikliklerin sınıflandırma performansı üzerindeki etkisi test edilmiştir. Yığın büyüklüğü değeri BERT için önerilen değerler içerisinde tutulmuştur. Ancak bu çalışma özelinde değerlendirilecek olursak sınıflandırmalar 16, 32, 64, 128 değerleri kullanılarak yapıldığında, yığın büyüklüğü arttıkça sınıflandırma performansının azaldığı gözlenmektedir. Yığın büyüklüğünün artması modelin eğitime hızını artırır. Ancak bu durum da güncelleme adımları daha az sayıda yapılır. Sadece yığın büyüklüğü değeri sınıflandırma performansına tek başına etki edemez. Girdi çeşidi, kurulan model, optimizasyon algoritması gibi değerler de sınıflandırma performansını etkiler.

**Tablo 4.** Farklı yığın büyüklüğü değerlerine göre sınıflandırma performansı (Classification performance according to different batch size values)

| Öğrenme Oranı | Yığın Büyüklüğü | Kesinlik k | Duyarlılık k | F1   | Doğruluk k |
|---------------|-----------------|------------|--------------|------|------------|
| lr=1e-05      | 16              | 0,55       | 0,55         | 0,54 | 0,55       |
|               | 32              | 0,52       | 0,53         | 0,51 | 0,53       |
|               | 64              | 0,53       | 0,53         | 0,51 | 0,53       |
|               | 128             | 0,49       | 0,51         | 0,48 | 0,51       |

Girdi=İstem, Döngü sayısı=5

#### 4.2. Farklı Öğrenme Oranlarına Göre Sınıflandırma Performansındaki Değişim

(The Change in Classification Performance According to the Different Learning Rates)

Sınıflandırma performansını en çok etkileyen hiper parametrelerden biri de öğrenme oranındaki değişimdir. Öğrenme oranı sınıflandırma problemlerinde model ağırlıklarının eğitim sırasında ne kadar güncellendiğini belirten bir parametredir. Tablo 5'deki veriler incelendiğinde öğrenme oranındaki azalma genel olarak sınıflandırma doğruluğundaki artışla ters orantılı olarak çalışmaktadır. Bu durumda

öğrenme oranı küçüldükçe genelde elde edilen doğruluk değerleri artmaktadır. Ancak öğrenme oranını da probleme göre ayarlamak da önemlidir. Örneğin 2e-06 gibi çok küçük öğrenme oranı kullanmak modelin sınıflandırma performansını çok düşürmektedir. Standart bir değer bu parametrede de geçerli değildir.

Patent metinlerinin analizinin gerçekleştirildiği bu çalışmada (Tablo 5'de gösterilmiştir) döngü değerleri karşılaştırıldığında döngü değerinin artmasının doğruluk değerine olumlu katkı sağladığı gözlemlenmemiştir. Döngü değeri 1'den 2'ye çıktığında sınıflandırma performansında artış olmaktadır. Ancak bu artış oranı döngü değerleri arttıkça yerini sabit kalmaya bırakmaktadır. Ayrıca, döngü sayısının tahmin doğruluğuna etkisi BERT kullanılan çalışmalarda belirgin bir şekilde daha yüksek olsaydı, 10 deneme ile yapılan analizlerde çok daha yüksek bir doğruluk değerinin bulunması gerekirdi. Dolayısıyla 5 döngüden sonra döngü değerinin artırılması tahmin doğruluğunu arttırmamaktadır. Bu test, literatürde de benzer açıklamaları teyit etmek amacıyla gerçekleştirilmiştir [58, 69]. Şekil 5 farklı öğrenme oranlarının farklı döngü sayılarındaki değişimini göstermektedir.

**Tablo 5:** Farklı öğrenme oranlarına göre sınıflandırma performansı (The classification performance according to different learning range)

| Döngü Sayısı | Öğrenme oranı | Kesinlik | Duyarlılık | F1   | Doğruluk |
|--------------|---------------|----------|------------|------|----------|
| 5            | 2e-06         | 0,26     | 0,34       | 0,26 | 0,34     |
|              | 1e-05         | 0,49     | 0,51       | 0,48 | 0,51     |
|              | 2e-05         | 0,54     | 0,54       | 0,53 | 0,54     |
|              | 3e-05         | 0,55     | 0,55       | 0,54 | 0,55     |
|              | 4e-05         | 0,55     | 0,54       | 0,53 | 0,54     |
| 10           | 5e-05         | 0,55     | 0,54       | 0,53 | 0,54     |
|              | 2e-06         | 0,2      | 0,29       | 0,21 | 0,21     |
|              | 1e-05         | 0,55     | 0,54       | 0,54 | 0,54     |
|              | 2e-05         | 0,55     | 0,54       | 0,54 | 0,54     |
|              | 3e-05         | 0,55     | 0,54       | 0,54 | 0,54     |
|              | 4e-05         | 0,55     | 0,53       | 0,53 | 0,53     |
|              | 5e-05         | 0,54     | 0,53       | 0,52 | 0,53     |

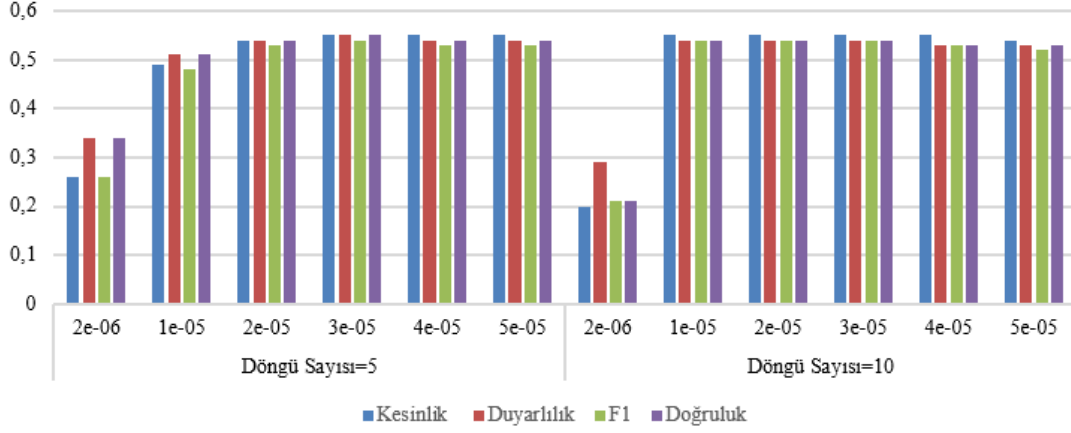
Girdi=Özet, Yığın Büyüklüğü= 64

#### 4.3. Farklı Patent Metni Bölümlerine Göre Sınıflandırma Performansındaki Değişiklik (The Change in Classification Performance According to Different Patent Text Parts)

Patent dokümanının farklı metin bölümleri ve bu bölümlerin çeşitli kombinasyonlar ile yapılan çalışma sonuçları Tablo 6'da gösterilmiştir. Bölümler tek başına incelendiğinde en uzun metinsel içeriğe sahip bölüm daha yüksek doğrulukta sınıflandırma performansına ulaşmıştır. Bunun nedeninin başlık ve özet gibi kısa bölümündeki kelime ve cümle sayısının istem ve tarifname gibi nispeten daha uzun bölümüne göre daha az olması gösterilebilir. Çünkü Bert algoritması yapısı gereği sıralı cümlelere ve o cümlelerin içindeki kelimelerin konumuna göre sınıflandırma performansını gerçekleştiriyor. Dolayısıyla sınırlı sayıda cümle içeren ve istemlere göre daha kısa olan özet bölümünde tahmin doğruluğunun daha düşük çıkması Bert algoritmasının çalışma performansına uyumludur.

Çeşitli kombinasyonların (Başlık+ Özet, Özet+İstem, Başlık+İstem, Başlık+Özet+İstem,) girdi olarak kullanıldığı sınıflandırma problemlerinde de (tarifname kısmı hariç) bu bölümlerin tek başına girdi olarak kullanımına göre daha yüksek doğruluğa ulaşılmıştır. Tam metin (başlık + özet + istem+ tarifname) verisinin girdi olarak alındığı sınıflandırma %58 ile en yüksek doğruluğa ulaşmıştır. Bu model ile ulaşılan doğruluk değeri literatürde Abdelgawad vd. [70] sonucundan (%55,2) daha yüksek olan %58 ile daha yüksek sınıflandırma sonucuna ulaşmıştır.

### Öğrenme Oranlarının sınıflandırma performansına etkisi



Şekil 5. Farklı öğrenme oranlarının sınıflandırma performansına olan etkisi  
(The effect of different learning rates on classification performance)

Tablo 6. Farklı patent metin bölümleri için sınıflandırma performansı  
(Classification performance for different patent text parts)

| Girdi             | Kesinlik | Duyarlılık | F1   | Doğruluk |
|-------------------|----------|------------|------|----------|
| Başlık            | 0,47     | 0,48       | 0,46 | 0,49     |
| Özet              | 0,54     | 0,54       | 0,53 | 0,54     |
| İstem             | 0,55     | 0,54       | 0,53 | 0,54     |
| Tarifname         | 0,57     | 0,57       | 0,56 | 0,57     |
| Tam metin         | 0,58     | 0,58       | 0,56 | 0,58     |
| Başlık+ Özet      | 0,55     | 0,56       | 0,54 | 0,56     |
| Özet+İstem        | 0,55     | 0,55       | 0,54 | 0,55     |
| Başlık+İstem      | 0,57     | 0,56       | 0,55 | 0,56     |
| Başlık+Özet+İstem | 0,55     | 0,55       | 0,54 | 0,55     |

Öğrenme Oranı=2e-05, Yığın boyutu=64, Döngü sayısı=5

### 5. Sonuç ve Gelecek Çalışma (Conclusion and Future Work)

Patent sınıflandırması, patent başvurularının yapıldığı teknik alanları belirlemek için kullanılan bir sınıflandırma sistemidir. Bu sistem sayesinde patent başvuruları daha kolay bir şekilde araştırılabilir ve değerlendirilebilir. Günümüzde yeni buluşların sistemde kaydının tutulması ve ilgililerin bu alanda güncellenmesi açısından sınıflandırma sistemleri sürekli olarak bu tempoya adapte olmaya çalışıyor. Derin öğrenme alanındaki çalışmalar patent sınıflandırma problemlerine de çözüm bulma konusunda ilerlemektedir. Derin öğrenme dünyasına 2018 yılında katılan BERT modeli, patent sınıflandırma alanında yapılan çalışmalarda da yer almaya devam etmektedir.

Bu çalışmada BERT tabanlı model kullanılarak bir patent sınıflandırma çalışması gerçekleştirilmiştir. Çalışmada patent metinlerinin farklı bölümlerinin yanı sıra farklı hiper parametrik değişkenler ile sınıflandırma performansı ölçülmeye çalışılmıştır. Ulaşılan en yüksek doğruluk değeri (%58) "tarifnamenin" girdi olarak kullanıldığı, öğrenme oranının "2e-05" olduğu ve yığın büyüklüğünün "64" olarak alındığı modelde gerçekleşmiştir. Bu doğruluk değeri başka birçok patent sınıflandırma problemi için düşük olarak görülse de Wipo-alpha veri seti özelinde, literatürdeki benzer çalışmalar ile yarışabilecek düzeydedir ve bu anlamda literatüre katkı sağlamaktadır. Bu çalışmada, hiperparametre analizinin önemine vurgu yapılmıştır. Kullanılan parametrelerin detaylı olarak sunulması, literatüre katkı sağlamakla kalmayıp, aynı zamanda benzer çalışmalar yapan araştırmacılara da yol göstermektedir. Çünkü bu tarz

hiperparametre analizleri, herhangi bir makalenin ön araştırma aşamasında analiz için araştırmacının ihtiyacı olan bilgileri sağlamaktadır. Bu veri setinde bazı sınıfların diğerlerine göre daha fazla veya daha az örnek sayısına sahip olması, sınıflandırma modelinin bazı sınıfları tanımda daha iyi (başarılı) veya kötü (başarısız) olmasına neden olabilir. Bu durum sınıflandırma doğruluğunun düşmesine sebep olabilir. Kullanılan dört farklı yığın büyüklüğü, altı farklı öğrenme oranı ile yapılan denemeler ve patent belgesinin dört farklı metin bölümünün ve çeşitli kombinasyonlarının girdi verisi olarak modele eklenmesinin oluşturduğu geniş hiperparametre analizi ile bu çalışma literatüre katkı sağlamaktadır. Bu çalışmada, hiperparametre analizinin önemine vurgu yapılmıştır. Kullanılan parametrelerin detaylı olarak sunulması, literatüre katkı sağlamakla kalmayıp, aynı zamanda benzer çalışmalar yapan araştırmacılara da yol göstermektedir. Çünkü bu tarz hiperparametre analizleri, herhangi bir makalenin ön araştırma aşamasında analiz için araştırmacının ihtiyacı olan bilgileri sağlamaktadır. Bu veri setinde bazı sınıfların diğerlerine göre daha fazla veya daha az örnek sayısına sahip olması, sınıflandırma modelinin bazı sınıfları daha iyi veya daha kötü tanımasına neden olmuş olabilir. Bu durum sınıflandırma doğruluğunun düşmesine sebep olabilir. Hiperparametre analizlerinde öğrenme oranı ve döngü sayısı açısından literatürdeki çalışmalarla benzer sonuçlara ulaşılmışken (4-5 döngü sayısının yeterli olduğu ve döngü sayısını çok arttırmanın sınıflandırma performansına anlamlı bir katkı sağlamadığı görülmüştür.), yığın büyüklüğü konusunda literatürden farklı sonuçlar elde edilmiştir. Bu durumun alınan yığın büyüklüğü değerlerinin daha düşük tutularak deneylerin tekrarlanmasıyla ileri de yapılacak çalışmalar için önem arz edeceği düşünülmektedir. Patent metinlerinin seçimi açısından literatürde patent özetinin veya patent istem bölümünün yeterli olduğu sonucuna ulaşan birçok çalışma olmasına rağmen, bu çalışmadan elde edilen sonuçlara göre; sadece özet ya da sadece istem bölümünü kullanmak yeterli olmamaktadır. Patent belgesinin tüm metinsel içeriklerini kullanmak daha yüksek sınıflandırma doğruluğuna ulaşılmasını sağlamıştır. Ancak bu durum yüksek hesaplama zamanına sebep olmaktadır. Bu çalışmada bir teknik sorun ile de karşılaşmıştır. Google Colab ortamında gerçekleştirilen çalışma sınırlı sayıda Ram kullanımı ve Gpu değeri gibi kısıtlamalar bazı modellerde çalışmayı sekteye uğratmıştır.

Gelecekte BERT-large ve diğer BERT ailesi üyeleri ile sınıflandırma işleminin denenmesi arzu edilmektedir. BERT üzerine yapılan çalışmalarda kelime gömme kullanılması gerekli olmamakla birlikte

bazı çalışmalarda bu durum göz ardı edilerek kelime gömme kullanılabilir. İleride patent metinlerinden türetilen ve içerisinde hukuki terimlerden teknolojik terimlere birçok patente özgü kelimelerin bulunduğu spesifik bir kelime gömme algoritması ile sınıflandırma yapılması arzu edilmektedir.

## Kaynaklar (References)

- Asche, G., "80% of technical information found only in patents" – Is there proof of this ?, *World Patent Information*, 48, 16–28, 2017.
- Seo, W. A patent-based approach to identifying potential technology opportunities realizable from a firm's internal capabilities, *Computers & Industrial Engineering (Comput. Ind. Eng.)*, 171, 108395, 2022.
- Sofean, M., Deep learning based pipeline with multichannel inputs for patent classification, *World Patent Information*, 66, 102060, 2021.
- Giordano V., Chiarello F., Melluso N., Fantoni G. and Bonaccorsi A., Text and Dynamic Network Analysis for Measuring Technological Convergence: A Case Study on Defense Patent Data, *IEEE Transactions on Engineering Management (IEEE Trans. Eng. Manage)*, 70, 4, 1490-1503, 2023.
- Durmusoglu, A., Updating technology forecasting models using statistical control charts, *Kybernetes*, 47 (4), 672–688, 2018.
- Hain, D.S., Jurowetzki, R., Buchmann, T., and Wolf, P. A., Text-Embedding-based Approach to Measure Patent-to-Patent Technological Similarity, *Technological Forecasting and Social Change (Technol. Forecasting Social Change)*, 177, 121559, 2022.
- Yan, B., and Luo, J., Measuring technological distance for patent mapping, *The Journal of the Association for Information Science and Technology*, 68 (2), 423–437, 2017.
- Erzurumlu, S.S., and Pachamanova, D., Topic modeling and technology forecasting for assessing the commercial viability of healthcare innovations, *Technological Forecasting and Social Change (Technol. Forecasting Social Change)*, 156, 120041, 2020.
- Grawe, M.F., Martins, C.A., and Bonfante, A.G., Automated patent classification using word embedding, *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017, Cancun, Mexico, 408–411, 18-21 Aralık 2017*.
- Altuntas, S., Dereli, T., and Kusiak, A., Analysis of patent documents with weighted association rules, *Technological Forecasting and Social Change (Technol. Forecasting Social Change)*, 92, 249–262, 2015.
- Mutlu N.G., Altuntaş S., Monitoring technological changes with statistical control charts based on patent data, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 36 (4), 1875-1892, 2021.
- Li, S., Hu, J., Cui, Y., and Hu, J., DeepPatent: patent classification with convolutional neural networks and word embedding, *Scientometrics*, 117, 721–744, 2018.
- Yücesoy Kahraman, S., Dereli, T., Durmuşoğlu, A. Forty years of automated patent classification, *International Journal of Information Technology & Decision Making (Int. J. Inf. Technol. Decis. Making)*, 2023. <https://doi.org/10.1142/S0219622023500165>
- Fall, C.J., Törösvári, A., Fiévet, P., and Karetka, G., Automated categorization of German-language patent documents, *Expert Systems with Applications (Expert Syst. Appl.)*, 26 (2), 269–277, 2004.
- Fall, C., and Benzineb, K., Literature survey: Issues to be considered in the automatic classification of patents, *World Intellectual Property Organization*, 2002.
- Piroi, F., and Hanbury, A., Multilingual Patent Text Retrieval Evaluation: CLEF-IP, *Information Retrieval Evaluation in a Changing World, The Information Retrieval Series*, Springer Nature Switzerland AG, 41, 365–387, 2019.
- Roda, G., Tait, J., Piroi, F., and Zenz, V., CLEF-IP 2009: Retrieval Experiments in the Intellectual Property Domain, *Multilingual Information Access Evaluation I. Text Retrieval Experiments CLEF 2009, Lecture Notes in Computer Science*, 6241. Springer, Berlin, Heidelberg, 2010.
- Karaca A., Aydın Ö., Generating headlines for Turkish news texts with transformer architecture based deep learning method, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 39 (1), 485–496, 2023.
- Lee, J.S., and Hsiang, J., Patent classification by fine-tuning BERT language model, *World Patent Information*, 61, 101965, 2020.
- Bekamiri, H., Hain, D.S., and Jurowetzki, R., PatentSBERTa: A Deep NLP based Hybrid Model for Patent Distance and Classification using Augmented SBERT, <https://arxiv.org/abs/2103.11933>, Yayın tarihi Mar 22 2021, Güncelleme tarihi Mayıs 29 2021, Erişim tarihi Temmuz 25, 2023.
- Chen, Y.L., and Chang, Y.C. A three-phase method for patent classification, *Information Processing & Management (Inf. Process. Manage)*, 48 (6), 1017–1030, 2012.
- Yücesoy, S., Dereli, T., and Durmuşoğlu, A., Patent Classification via Textual Analysis Which Sections to be Included? 2018 International Conference on Artificial Intelligence and Data Processing, IDAP 2018, Malatya, Turkey, 1-8, 28-30 Eylül 2018.
- Trappey, A., Development of a patent document classification and search platform using a back-propagation network, *Expert Systems with Applications (Expert Syst. Appl.)*, 31, (4), 755-765, 2006.
- Benzineb, K., and Guyot, J., Automated Patent Classification, *Current challenges in patent information retrieval, The Information Retrieval Series 29, Springer-Verlag Berlin Heidelberg*, 239-261, 2011.
- Rousu J., Saunders J, Szedmak, C., Shawe S., Taylor, J., Bennett, K.P., and Parrado-Hernández, E., Kernel-Based Learning of Hierarchical Multilabel Classification Models, *Journal of Machine Learning Research*, 7, 1601–1626, 2006.
- Wu, C., Ken, Y., Huang, T, Patent classification system using a new hybrid genetic algorithm support vector machine, *Applied Soft Computing* 10, 4, 1164-1177, 2010.
- Hu, J., Li, S., Hu, J., and Yang, G., A hierarchical feature extraction model for multi-label mechanical patent classification, *Sustainability*, 10 (1), 1-22, 2018.
- Alla G. Kravets, Nikita Lebedev, and Maxim Legenchenko, Patents images retrieval and convolutional neural network training dataset quality improvement, *IV International Research Conference Information Technologies in Science, Management, Social Sphere and Medicine ITSMSM 2017*, 287–293, 2017.
- Wang, Y., Du, J., Shao, Y., Li, A., and Xu, X., A Patent Text Classification Method Based on Phrase-Context Fusion Feature.", *Proceedings of 2021 Chinese Intelligent Automation Conference*, Deng Z., Springer, Singapore, 157-164, 2022.
- Kantar O., Kilimci Z.H., Deep learning based hybrid gold index (XAU/USD) direction forecast model, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 38 (2), 1117–1128, 2022.
- Sanh, V., Debut, L., Chaumond, J., and Wolf, T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, <https://arxiv.org/abs/1910.01108>, Yayın tarihi Ekim 2019, Güncelleme Tarihi, Mart 2020, Erişim Tarihi Temmuz 25, 2023.
- Devlin, J., Chang, M.-W., Lee, K., Google, K.T., and Language, A.I. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, <https://arxiv.org/abs/1810.04805>, Yayın tarihi Ekim 2018, Güncelleme Tarihi May 2019, Erişim Tarihi Temmuz 25, 2023.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., and Allen, P.G. (2019) RoBERTa: A Robustly Optimized BERT Pretraining Approach, <https://arxiv.org/abs/1907.11692>, Yayın Tarihi Temmuz 2019, Güncelleme Tarihi Temmuz 2019, Erişim Tarihi Temmuz 25, 2023.
- Li, S., Hu, J., Cui, Y., and Hu, J., DeepPatent: patent classification with convolutional neural networks and word embedding. *Scientometrics*, 117 (2), 721–744, 2018.
- Lim, S., and Kwon, Y., IPC multi-label classification applying the characteristics of patent documents, *Advances in Computer Science and Ubiquitous Computing, UCAWSN CUTE CSA 2016, Park, J., Lecture Notes in Electrical Engineering*, Springer, Singapore, 421, 166–172, 2017.
- Hu, J., Li, S., Hu, J., and Yang, G., A Hierarchical Feature Extraction Model for Multi-Label Mechanical Patent Classification, *Sustainability*, 10 (1), 219,2018.
- Lee, J.S., and Hsiang, J., Patent classification by fine-tuning BERT language model, *World Patent Information*, 61, 101965, 2020.
- Fall, C.J., Törösvári, A., Benzineb, K., and Karetka, G., Automated categorization in the international patent classification, *ACM SIGIR Forum*, 37 (1), 10–25, 2003.
- Qiu, X., Huang, X.-J., Liu, Z., and Zhou, J., Hierarchical Text Classification with Latent Concepts, *Proceedings of the 49<sup>th</sup> Annual Meeting of the Association for Computational Linguistics: short papers 598–602, Portland, Oregon, 19-24 Haziran 2011*.

40. Seneviratne, D., Geva, S., Zuccon, G., Ferraro, G., and Meireles, M., Linking patents to knowledge sources: A context matching technique using automatic patent classification, *Proceedings of the 23rd Australasian Document Computing Symposium*, 1-4, 2018.
41. Aiolli, F., Cardin, R., Sebastiani, F., and Sperduti, A., Preferential Text Classification: Learning Algorithms and Evaluation Measures, *Information retrieval*, 12, 559–580, 2009.
42. Chen, Y.L., Chang, Y.C., A three-phase method for patent classification. *Information Processing & Management (Inf. Process. Manage)*, 48 (6), 1017–1030, 2021.
43. Altuntas, F., Gök, M.Ş., Analysis of patent documents with utility mining: a case study of wind energy technology. *Kybernetes*, 50 (9), 2548–2582, 2020.
44. Perez-Molina, E. (2018) The role of patent citations as a footprint of technology. *Journal of the Association for Information Science and Technology (J. Assoc. Inf. Sci. Technol)*, 69 (4), 610–618, 2018.
45. Minin, V.A., Zatsman, I.M., Khavanskov, V.A., and Shubnikov, S.K., Methods of indicator-based assessment of knowledge transfer from science to technology, *Automatic Documentation and Mathematical Linguistics (Autom. Doc. Math. Ling.)*, 51 (4), 180–190, 2017.
46. Risch, J., and Krestel, R., Learning patent speak: Investigating domain-specific word embeddings, *13th International Conference on Digital Information Management, ICDIM 2018, Berlin, Germany*, 63–68, 24-26 Eylül 2018.
47. Sun, C., Qiu, X., Xu, Y., and Huang, X., How to Fine-Tune BERT for Text Classification? *Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China*, 194–206, 18-20 Ekim 2019.
48. Roudsari A.H., Jafar Afshar Lee, Wookey Lee S., PatentNet: multi-label classification of patent documents using deep learning based language understanding, *Scientometrics*, 127, 207–231, 2022
49. Wang, J., Zhang, Z., Feng, L., Lin, K.Y., and Liu, P., Development of technology opportunity analysis based on technology landscape by extending technology elements with BERT and TRIZ. *Technological Forecasting and Social Change (Technol. Forecasting Social Change)*, 191, 122481, 2023.
50. Qiu, Z., and Wang, Z., What is your next invention? - A framework of mining technological development rules and assisting in designing new technologies based on BERT as well as patent citations, *Computers in Industry (Comput Ind)*, 145, 103829, 2023.
51. Choi, S., Lee, H., Park, E., and Choi, S., Deep learning for patent landscaping using transformer and graph embedding, *Technological Forecasting and Social Change (Technol. Forecasting Social Change)*, 175, 121413, 2022.
52. Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I., Attention Is All You Need, *Advances in neural information processing systems*, 31<sup>st</sup> Conference on Neural Information Processing Systems (NIPS2017), Long Beach, CA, USA.1-11, 4-9 Aralık 2017.
53. Golubev, A.A., and Loukachevitch, N. V., Use of Bert Neural Network Models for Sentiment Analysis in Russian. *Automatic Documentation and Mathematical Linguistics*, 55 (1), 17–25, 2021.
54. Syed, M.H., and Chung, S.-T., MenuNER: Domain-Adapted BERT Based NER Approach for a Domain with Limited Dataset and Its Application to Food Menu Domain, 2021.
55. Ozan, Ş., Özdil, U., Taşar, D.E., Arslan, B., and Polat, G., BERT Modeli'nin Sınıflandırma Doğruluğunun Sıfır-Atış Öğrenmesi ile Artırılması. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 14 (2), 99–108, 2021.
56. Sevlı, O., and Kemalöglu, N. (2021) Classification of Tweets about Disasters as Real or Unreal by Google BERT Model Abstract, *Veri Bilim Dergisi*, 4 (1), 31-37, 2021.
57. Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., and Dean, J. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, <https://arxiv.org/abs/1609.08144>, Yayın tarihi Eylül 26 2016, Güncelleme Tarihi Ekim 8 2016, Erişim Tarihi Temmuz 25, 2023.
58. Jiang, S., Bengue, C., and King, W.C., BERTVision --A Parameter-Efficient Approach for Question Answering, <https://arxiv.org/abs/1609.08144>, Yayın tarihi Şubat 24 2022, Erişim Tarihi Temmuz 25, 2023, 2022
59. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R., and Research, G. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations, <https://arxiv.org/abs/1909.11942>, Yayın tarihi Eylül 26 2019, Güncelleme Tarihi Şubat 9 2020, Erişim Tarihi Temmuz 25, 2023.
60. Martin, L., Muller, B., Ortiz Suárez, P.J., Dupont, Y., Romary, L., Villemonte De La Clergerie, É., Seddah, D., and Sagot, B. CamemBERT: a Tasty French Language Model, <https://arxiv.org/abs/1911.03894>, Yayın tarihi Kasım 10 2019, Güncelleme Tarihi Mayıs 21 2020, Erişim Tarihi Temmuz 25, 2023.
61. Gao, Z., Feng, A., Song, X., and Wu, X., Target-dependent sentiment classification with BERT. *IEEE Access*, 7, 154290–154299, 2019.
62. The World Intellectual Property Organization (WIPO) International Classifications - IT Support Area, <https://www.wipo.int/classifications/ipc/en/ITsupport/Version20210101/transformations/stats>, Yayın tarihi Ocak 1 2015, Güncelleme Tarihi Aralık 27 2022, Erişim Tarihi Mayıs 3, 2023.
63. Aroyehun, S.T., Angel, J., Majumder, N., Gelbukh, A., and Hussain, A. (2021) Leveraging label hierarchy using transfer and multi-task learning: A case study on patent classification, *Neurocomputing*, 464, 421–431, 2021.
64. Smith, L.N., Cyclical learning rates for training neural networks. *Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017, Santa Rosa, CA, USA* 464–472, 24-31 Mart, 2017.
65. Popel, M., and Bojar, O., Training Tips for the Transformer Model. *The Prague Bulletin of Mathematical Linguistics*, 110 (1), 43–70, 2018.
66. Tan, K.L., Lee, C.P., Anbananthen, K.S.M., and Lim, K.M., RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis with Transformer and Recurrent Neural Network. *IEEE Access*, 10, 21517–21525, 2022.
67. Huang, W.C., Wu, C.H., Luo, S.B., Chen, K.Y., Wang, H.M., and Toda, T., Speech recognition by simply fine-tuning BERT. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, Toronto, ON, Canada*, 7343–734706, 06-11 Haziran, 2021.
68. Bozkur, R., Cicioğlu, M., Toğay, C., and Çalhan, A., The Effect of Flow Times on Machine Learning Algorithms in Flow Based Classification of Network Traffic, *European Journal of Science and Technology Special Issue*, 36, 276–283, 2022.
69. Hao, Y., Dong, L., Wei, F., Xu, K., Hao, Y., Dong, L., Wei, F., and Xu, K., Visualizing and Understanding the Effectiveness of BERT, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9<sup>th</sup> International Joint Conference on Natural Language Processing*, 4143–4152 Hong Kong, China, 3–7 Kasım, 2019.
70. Abdelgawad, L., Kluegl, P., Genc, E., Falkner, S., and Hutter, F. Optimizing Neural Networks for Patent Classification, *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 688-703, Würzburg, Germany, 16–20 Eylül, 2019.