



IX. Ulaştırma ve Lojistik Ulusal Kongresi

24-25 Ekim 2025 | Çanakkale Onsekiz Mart Üniversitesi

<https://www.ulk.ist/>



Lojistik Sektöründe Akıllı Doküman Kategorileme

Selçuk Zereyalp

YL Öğrencisi,

Sertrans Logistics

selcuk.zereyalp@sertrans.com.tr

Kemal Soğukçeşme

Sertrans Logistics

kemal.sogukcesme@sertrans.com.tr

Bahadır Fatih Yıldırım

İstanbul Üniversitesi

bahadirf.yildirim@istanbul.edu.tr

Özet

Çalışmada, geleneksel belge sınıflandırma yöntemlerinin ötesine geçerek yapay zeka tabanlı, özelleştirilebilir ve yüksek doğruluk oranına sahip bir belge sınıflandırma altyapısı sunmayı hedeflemektedir. Günümüzde birçok sektörde doküman yönetimi manuel veya kurala dayalı sistemlerle gerçekleştirilmektedir. Bu yaklaşım hem zaman kaybına hem de insan hatasına açıktır. Projemiz ise, derin öğrenme ve doğal dil işleme (NLP) tekniklerini bir araya getirerek lojistik sektörde kullanılmakta olan çeşitli belgeleri içeriklerine göre otomatik ve anlam odaklı şekilde sınıflandıracaktır. Sınıflandırılacak Doküman Türleri Sistem, lojistik sektörde sıklıkla kullanılan temel doküman türlerini tanıyacak ve otomatik olarak sınıflandıracaktır. Bunlar arasında: • ATR Belgesi: Avrupa Birliği ile Türkiye arasındaki ticarete gümrük vergisi muafiyeti sağlayan belgelerden biridir. Sistemin bu belgeyi doğru tanıması, dış ticaret süreçlerinin hızlanmasını ve beyan hatalarının önlenmesini sağlar. • CMR (Taşıma Senedi): Uluslararası karayolu taşımacılığında zorunlu olan taşıma belgesidir. Bu belgenin doğru sınıflandırılması, taşıma takibi, yük sigortası ve hukuki süreçlerin dijital yönetimi açısından kritik öneme sahiptir. • Commercial Invoice (Ticari Fatura): İhracat işlemlerinde mal bedeli ve satış koşullarını belirten temel belgedir. Otomatik sınıflandırma, ticari verilerin muhasebe ve gümrük sistemleriyle entegrasyonunu kolaylaştırır. • Proforma Invoice: Satış öncesinde hazırlanan teklif niteliğindeki faturadır. Bu belgelerin tanınması, teklif ve sipariş süreçlerinin dijitalleştirilmesine katkı sağlar. • Fatura: Mal veya hizmet satışının yasal kayıdır. Faturaların otomatik olarak ayrıştırılması, e-fatura sistemleriyle entegrasyonu ve denetim süreçlerinde zaman kazandırır. • Gümrük Beyannamesi: İthalat ve ihracat işlemlerinde zorunlu bir evraktır. Bu belgelerin doğru kategorize edilmesi, beyan süreçlerinde hata payını azaltır ve yasal uyumluluk sağlar. • T1 / T2 Transit Belgeleri: Malların bir ülkeden diğerine geçişinde kullanılan transit belgeleridir. Bu belgelerin otomatik olarak tanımlanması, taşıma zincirindeki duraklar arasında bilgi akışını hızlandırır. • TR Belgeleri: İç taşımacılıkta veya Türkiye sınırları içerisindeki hareketlerde kullanılan ulusal taşıma belgeleridir. Bu belgelerin dijital yönetimi, operasyonel verimliliği artırır. Bu dokümanların tamamı sistem tarafından otomatik olarak tanınacak, sınıflandırılacak ve arşivlenecektir. Sistem Bileşenleri Belge Girdisi Katmanı: PDF, Word, e-posta, düz metin gibi farklı formatlardaki belgeleri sisteme alır. Ön İşleme Modülü: Belgelerden çıkarılan metinler temizlenir (tokenization, stopword removal, lemmatization vb.) ve sınıflandırmaya uygun hale getirilir. Gerekliğinde dil algılama ve çok dilli içerik



yönetimi yapılır. Özellik Çıkarım (Feature Extraction): Metinlerden TF-IDF, word embeddings (Word2Vec, GloVe, BERT) gibi yöntemlerle anlamlı vektör temsilleri üretilir. Sınıflandırma Motoru: Belirlenen kategorilere göre denetimli makine öğrenmesi algoritmaları veya derin öğrenme modelleri (örneğin BERT tabanlı transformer modelleri) kullanılarak belge sınıflandırması yapılır. Veritabanı ve Arşivleme Sistemi: Sınıflandırılmış belgeler, kategori, tarih, dosya tipi gibi metadata'larla birlikte indekslenerek veritabanına kaydedilir. Temel Tasarım İlkeleri

- Otomasyon: Belge sınıflandırma süreci tamamen otomatiktir, kullanıcı müdahalesi gerektirmez.
- Esneklik ve Öğrenebilirlik: Yeni belge türleri eklendikçe sistem, sürekli olarak eğitilerek gelişen yapay zeka modeli ile daha doğru sonuçlar üretir.
- Çok Dilli Destek: İngilizce, Türkçe ve diğer dillerdeki belgelerle çalışabilir.
- Gizlilik ve Güvenlik: Belgeler şifreli ortamda işlenir; veri güvenliği ve gizliliğe öncelik verilir.

Yenilikçi Yönler

- Aktif öğrenme entegrasyonu: Sistem, kullanıcı geri bildirimlerine göre zamanla kendi sınıflandırma algoritmalarını iyileştirme kapasitesine sahip olacak şekilde tasarlanacaktır.
- Sektöre özgü özelleştirme: Uluslararası lojistik süreçlerine özel (ATR, CMR, T1, T2 vb.) modeller geliştirilecektir.
- Akıllı belge tanıma: OCR ve NLP birlikte kullanılarak, belge taramaları üzerinden içerik tabanlı sınıflandırma yapılabilecektir.
- Ar-Ge Niteliği
- Yeni algoritmaların denenmesi: Transformer tabanlı modeller kullanılarak mevcut yöntemlerin ötesinde doğruluk oranları hedeflenmektedir.
- Veri kümesi oluşturulması: Türkçe veya sektör özelinde sınıflandırma için kullanılacak etiketli veri setleri oluşturulmakta, bu da akademik ve endüstriyel katkı sağlayacaktır.
- Model optimizasyonu: Geliştirilen modeller düşük işlem gücüne sahip cihazlarda bile çalışacak şekilde optimize edilecektir.
- Akademik katkı potansiyeli: Elde edilen bulgular akademik yayınlara dönüştürülerek alana katkı sağlayacaktır.

Kullanım Senaryosu (Örnek) Bir kullanıcı, lojistik operasyonlarına ait bir belgeyi sisteme yüklediğinde: 1. Sistem belgeyi tarar, içeriğini çıkarır. 2. Ön işleme ve özellik çıkarımı yaparak belgeyi "CMR", "ATR", "Fatura", "Gümrük Beyannamesi" gibi kategorilerden birine sınıflandırır. 3. Uygun klasöre kaydeder ve meta verileri (tarih, gönderici, belge türü vb.) ile birlikte arşivler. Bu sayede belge yönetimi süreci hızlanır, insan hatası ortadan kalkar ve lojistik süreçlerin dijital dönüşümü desteklenir.

Anahtar Kelimeler:

Yapay zeka, Derin öğrenme, Doğal dil işleme (NLP), Belge sınıflandırma, Doküman yönetimi sistemi, Otomatik belge analizi, Akıllı doküman yönetimi, Metin madenciliği, Veri işleme, Makine öğrenmesi, Lojistik sektörü, Dijital dönüşüm, Gümrük belgeleri, Taşıma belgeleri, Transit işlemler, Fatura yönetimi, Dış ticaret belgeleri, Operasyonel verimlilik, Lojistikte yapay zeka

Intelligent Document Classification in the Logistics Sector

Abstract

This paper aims to go beyond traditional document classification methods by providing an AI-based, customizable, and highly accurate document classification infrastructure. In many industries today, document management is performed manually or through rule-based systems. This approach is both time-consuming and prone to human error. Our project combines deep learning and natural language processing (NLP) techniques to automatically and semantically classify various types of documents commonly used in the logistics sector based on their content. Document Types to Be Classified The system will recognize and automatically classify the main types of documents frequently used in logistics operations, including: ATR Certificate: A document granting customs duty exemption for trade



between the European Union and Türkiye. Correct identification of this document speeds up foreign trade processes and prevents declaration errors. CMR (Consignment Note): A mandatory document in international road transport. Accurate classification of this document is crucial for transport tracking, cargo insurance, and the digital management of legal processes. Commercial Invoice: A key document that specifies the value of goods and sales terms in export operations. Automatic classification facilitates the integration of commercial data with accounting and customs systems. Proforma Invoice: A quotation-style invoice prepared before the sale. Recognizing these documents supports the digitalization of quotation and order management processes. Invoice: The legal record of a goods or services sale. Automatic extraction and categorization of invoices save time and enable seamless integration with e-invoicing and auditing systems. Customs Declaration: A mandatory document in import and export transactions. Correct categorization minimizes declaration errors and ensures regulatory compliance. T1 / T2 Transit Documents: Documents used for the movement of goods between countries. Their automatic identification accelerates information flow across different stages of the transport chain. TR Documents: National transport documents used for domestic logistics within Türkiye. Digital management of these documents improves operational efficiency. All of these document types will be automatically detected, classified, and archived by the system. System Components Document Input Layer: Accepts documents in various formats such as PDF, Word, email, or plain text. Preprocessing Module: Extracted text is cleaned (tokenization, stopword removal, lemmatization, etc.) and prepared for classification. Language detection and multilingual content management are also applied when needed. Feature Extraction: Meaningful vector representations are generated from text using techniques such as TF-IDF, Word2Vec, GloVe, or BERT. Classification Engine: Supervised machine learning or deep learning models (e.g., BERT-based transformer architectures) are used to classify documents into predefined categories. Database and Archiving System: Classified documents are indexed and stored in the database along with metadata such as category, date, and file type. Core Design Principles Automation: The entire document classification process is fully automated and requires no manual intervention. Flexibility and Learnability: As new document types are introduced, the system continuously learns and improves accuracy through retraining. Multilingual Support: The system can handle documents in English, Turkish, and other languages. Confidentiality and Security: Documents are processed in an encrypted environment with strict data privacy and protection measures. Innovative Features Active Learning Integration: The system will be capable of improving its classification algorithms over time based on user feedback. Sector-Specific Customization: Models will be tailored to international logistics workflows (e.g., ATR, CMR, T1, T2). Intelligent Document Recognition: By combining OCR and NLP, the system will perform content-based classification even on scanned or image-based documents. R&D Aspects Exploration of New Algorithms: Transformer-based models will be used to achieve higher accuracy rates compared to existing methods. Dataset Development: Labeled datasets specific to Turkish or logistics-related document classification will be created, contributing to both academic and industrial research. Model Optimization: Models will be optimized to run efficiently even on low-resource environments. Academic Contribution: Research findings will be published to support academic progress in AI and NLP fields. Use Case (Example) When a user uploads a logistics-related document to the system: The system scans and extracts the document's content. It preprocesses and extracts features, then classifies the document into one of the categories such as "CMR", "ATR", "Invoice",



or “Customs Declaration.” The document is stored in the corresponding folder along with metadata such as date, sender, and document type. This process accelerates document management, eliminates human error, and supports the digital transformation of logistics operations.

Keywords:

Artificial Intelligence, Deep Learning, Natural Language Processing (NLP), Document Classification, Document Management System, Automated Document Analysis, Intelligent Document Management, Text Mining, Data Processing, Machine Learning, Logistics Sector, Digital Transformation, Customs Documents, Transport Documents, Transit Operations, Invoice Management, Foreign Trade Documents, Operational Efficiency, Artificial Intelligence in Logistics

