



MAKİNE ÖĞRENME YARDIMI İLE ÇOK DİLLİ ADRES VERİTABANI İÇERİSİNDEN  
ÜLKE ADLARININ DERLENMESİ

Suat ATAN<sup>1</sup>

ÖZET

Tacirler, tabii oldukları mevzuat uyarınca dış ticaret işlemlerini herhangi bir ticaret odasına bildirmek durumundadırlar. Çalışma kapsamında bir ticaret odasına ait uygulama yazılımındaki 5 milyon kayıt incelenmiş ve ülke verilerine ait kolonda ülke bilgilerinin yer almadığı görülmüştür. Bu veri tabanında, yurtdışında bulunan İngilizce, Azerice, Almanca ve Türkçe firma adresleri bulunmakta adreslerde bulunan ülke adlarında yazım hataları, kısaltmalar ve yazım farklılıkları ya da ülke adının bulunmadığı haller söz konusudur. İlgili ticaret odası verilerini kategorize edememektedir. Bu çalışma kapsamında organizasyonlara ait eksik ve dağınık veri söz konusu olduğunda mevcut ve yapay verilerden üretilen makine öğrenme algoritmaları yardımı ile bu sorunun çözüldüğü gösterilmektedir. Multinomial Naïve Bayes algoritması yardımı ile öğrenen makine %82 doğruluk gücü ile adreslerdeki şehir adları ve benzeri yapılar yardımı ile tüm adresleri kodlamıştır. Bu sayede çoğunluğu ilgili ticaret odasına tabi tacirlerin en fazla dış ticaret gerçekleştirdiği ülke ve şehirler ortaya çıkarılmıştır. Algoritma geliştirilirken kullanılan esasların benzer problemlerin çözümünde bir alternatif olduğu değerlendirilmektedir.

**Anahtar Kelimeler:** İşletme, Yapay Zeka, Dijitalleşme

ICDBME

<sup>1</sup>Università degli Studi di Torino, suatatan@gmail.com ve ORCID: 0000-0003-3170-0969



**EXTRACTING COUNTRY NAMES FROM MULTILINGUAL POSTAL ADDRESS  
DATABASE WITH MACHINE LEARNING**

**Suat ATAN<sup>1</sup>**

**ABSTRACT**

Traders, are obliged to report their foreign trade transactions to any trade chamber of commerce in accordance with the legislation to which they are subject. Within the scope of the study, 5 million records in the software of a chamber of commerce were examined and it was seen that the country information was not included in the column of the country data. In this database, there are English, Azerbaijani, German and Turkish company addresses abroad, and there are spelling errors and spelling differences, abbreviations in the country names in the addresses, or cases where the country name is not available. The management of relevant chamber of commerce is unable to categorize its data. Within the scope of this study, it is shown that this problem is solved with the help of machine learning algorithms produced from existing and artificial data when it comes to missing and scattered data of organizations. The machine, which learned with the help of Multinomial Naïve Bayes algorithm, coded all addresses with the help of city names and similar structures in addresses with 82% accuracy. In this way, the countries and cities where the merchants, most of whom are subject to the relevant chamber of commerce, have the most foreign trade have been revealed. It is considered that the principles used while developing the algorithm are an alternative to solving similar problems.

**Keywords:** Business Administration, Artificial Intelligence, Digitalization

ICDBME

---

<sup>1</sup> Università degli Studi di Torino, suatatan@gmail.com ve ORCID: 0000-0003-3170-0969