Geliş/Received: 27.02.2024 | Kabul/Accepted: 31.08.2024 | Erken Görünüm/Early View: 16.09.2024

KOCATEPEİİBFD

Research Article (Special Issue) | Araştırma Makalesi (Özel Sayı)

An example of the application of artificial intelligence models in human resources processes

Selim Sürücü | Res. Assist., Çankırı Karatekin University, selimsurucu@karatekin.edu.tr, 0000-0002-8754-3846
Berk Küçük | Student, Çankırı Karatekin University, dev.berkkucukk@gmail.com, 0009-0005-3500-4540
Mustafa Kemal Aydın | Student, Çankırı Karatekin University, dev.mkaydin@gmail.com, 0009-0000-0306-9730

Corresponding author/Sorumlu yazar: Selim Sürücü selimsurucu@karatekin.edu.tr

Abstract

Creating job postings and selecting suitable candidates among these job postings is a challenging process. This process increases the workload of human resources and causes the process to proceed slowly. It is of great importance for human resources departments to utilize information processing technologies to create job postings effectively and to evaluate the CVs of applicants to these postings. This study introduces and analyzes two different technologies that can help human resources. In the process of preparing job advertisements in the field of IT, in the first stage, the word cloud method is used to decide which keywords should be emphasized in the advertisement texts. In the second stage, the resumes of the applicants are analyzed using three different deep learning models such as CNN (Convolutional Neural Network), GRU (Gated Recurrent Unit), and LSTM (Long Short-Term Memory) for classification purposes. While the performance of these models is evaluated using metrics such as accuracy, MCC, F_1 score, and MSE, the decision-making processes of the models with explainable artificial intelligence are also analyzed. In this context, the GRU model, which achieved an accuracy of 99%, provided the most superior result in this study and the literature. This research shows that deep learning models provide high accuracy rates and efficiency in human resources resume classification and candidate matching processes. It also explains that using the word cloud method, the most appropriate keywords can be identified, and advertisements can be created.

Keywords: Resume Classification, AI-Powered Human Resources Automation, Natural Language, Text Classification, Explainable Artificial Intelligence

JEL Codes: O15, M12, M15

İnsan kaynakları süreçlerinde yapay zekâ modellerinin uygulanmasına bir örnek

Öz

İş ilanı oluşturmak ve bu iş ilanları arasından uygun adayları seçmek zorlu bir süreçtir. Bu süreç insan kaynaklarının iş yükünü arttırmakta, sürecin yavaş ilerlemesine neden olmaktadır. İnsan kaynakları departmanlarının etkin bir şekilde iş ilanı oluşturabilmesi ve bu ilanlara yapılan başvuruların özgeçmişlerinin değerlendirilmesi süreçlerinde bilgi işlem teknolojilerinden yararlanılması büyük bir öneme sahiptir. Bu çalışma, insan kaynaklarına yardımcı olabilecek, iki farklı teknolojinin tanıtımını ve analizini yapmaktadır. Bilgi işlem teknolojileri alanında iş ilanlarının hazırlanması sürecinde, ilk aşamada kelime bulutu yöntemi kullanılarak ilan metinlerinde hangi anahtar kelimelerin vurgulanması gerektiğine karar verilir. İkinci aşamada, başvuran adayların özgeçmişleri, sınıflandırma amacıyla CNN (Convolutional Neural Network), GRU (Gated Recurrent Unit) ve LSTM (Long Short-Term Memory) gibi üç farklı derin öğrenme modeli kullanılarak analiz edilmiştir. Bu modellerin performansları, doğruluk, MCC, F_1 score ve MSE gibi metrikler kullanılarak değerlendirilirken, açıklanabilir yapay zekâ ile modellerin karar verme süreçleri de incelenmiştir. Bu çerçevede, %99'luk bir doğruluk başarısı sergileyen GRU modeli, bu çalışma kapsamında ve literatürde elde edilen en üstün sonucu sağlamıştır. Bu araştırma, derin öğrenme modellerinin, insan kaynakları alanında özgeçmiş sınıflandırma ve aday eşleştirme süreçlerinde yüksek doğruluk oranları ve verimlilik sağladığını göstermektedir. Ayrıca, kelime bulutu yöntemi kullanılarak en uygun anahtar kelimelerin belirlenerek ilanların oluşturulabileceğini de anlatmaktadır.

Anahtar Kelimeler: Özgeçmiş Sınıflandırma, Yapay Zekâ Destekli İnsan Kaynakları Otomasyonu, Doğal Dil İşleme, Metin Sınıflandırması, Açıklanabilir Yapay Zekâ

JEL Kodları: O15, M12, M15

Introduction

As one of the cornerstones of the modern business world, Human Resources (HR) plays a critical role in achieving an organization's strategic goals. Work in this field covers a wide range of areas, from employee recruitment, training and development,

How to cite this article / Bu makaleye atıf vermek için:
Sürücü, S., Küçük, B., & Aydın, M. K. (2024). An example of the application of artificial intelligence models in human resources processes. *KOCATEPEİİBFD, 26*(Özel Sayı), 101-116. https://doi.org/10.33707/akuiibfd.1443940

performance management, compensation, and benefits management (Mezhoudi et al., 2023). The main purpose of human resources is to effectively manage and develop human capital, the most valuable asset of an organization. It is a strategic investment to help employees achieve their career goals and increase the organization's overall productivity and competitiveness. Today, with technological advances and globalization, HR practices are constantly evolving, and organizations are faced with new challenges such as diversity, flexible working patterns, and increasing employee engagement. In this context, Human Resource management has become an indispensable strategic partner for successfully managing and future-proofing organizations.

Recruitment is one of human resources management's most critical and strategic activities. This process aims to bring the right talent to the company and involves identifying, assessing, and selecting candidates (Smith, 2021). A successful recruitment process starts with a detailed analysis of the positions needed, followed by preparing an effective job advertisement and selecting suitable candidates from among the applications. This process aims to find the most suitable person to meet the candidate's and organization's needs and expectations. Recruitment is not only about finding qualified individuals but also about selecting people who will fit into the company culture and contribute in the long term. Therefore, an effective recruitment process plays a vital role in the organization's success.

Today, with the digitalization of companies and the pandemic, the vast majority of recruitment processes (such as personnel advertisements and collection of resumes) take place online. In order to carry out the processes in a healthy, fast, and accurate manner, the human resources department needs to use technology more effectively (Poulose et al., 2024). For example, processes such as analyzing the personnel correctly in the recruitment advertisement, giving the right keywords, or examining and analyzing the collected resumes correctly must be carried out quickly and accurately. Effective use of technology can help human resources carry out these processes quickly and efficiently and make the right decisions.

With the increasing use of information technologies in companies, human resources processes have also adapted. With the developments in recent years, the use of Artificial Intelligence (AI) in automation systems is increasing. AI can be defined as the development of computer systems or algorithms that mimic human-like thinking, learning, decision-making, and problem-solving abilities. AI can be used for many different purposes in human resources processes. AI automates CV screening, candidate matching, and pre-assessment processes in recruitment, helping to identify the most suitable candidates quickly and objectively. By analyzing the experience, skills, and competencies of candidates, AI-supported tools select the most suitable candidates for open positions and accelerate the recruitment process. Another human resources process where AI is used is performance evaluation. Continuous monitoring and evaluation of employee performance can increase objectivity and precision in processes. In order to reduce the workload of human resources, chatbots, and virtual assistants can be utilized to improve the employee experience and increase employee engagement. When used in in-house training processes, it can offer personalized training programs in line with employees' individual learning styles and needs. Examining employees' performance data and learning progress, AI recommends the most appropriate training materials and modules. While traditional HR methods can be inefficient and subjective at times, AI-based solutions are reshaping the field by automating processes and making data-driven decisions. AI-supported human resources systems make human resources processes very fast while ensuring that the decisions made in the process are made with high accuracy. Companies with these systems have come to the forefront in their sectors.

The use of AI in the recruitment and selection stages of the human resources process has the potential to ensure efficiency and objectivity for both employers and candidates. AI technologies can quickly and accurately assess candidates' skills, experience, and suitability, thus improving and optimizing recruitment. By scanning candidates' resumes and application forms, AI can automatically identify the candidates who best match the job description and requirements. These systems can rank candidates based on criteria such as specific keywords, skills, education levels, and work experience and highlight the most suitable candidates. In addition, AI-supported human resources or resume analysis systems allow students to closely examine the pulse of the sectors and provide students with information about the features, such as technology and language, that they need to know according to the fields in which they want to work. While enabling students to plan their education accordingly, it also allows educational institutions to keep their curricula up-to-date.

This study was conducted to improve two human resources processes in information technology companies by using natural language processing methods, one of the AI fields. The first of these processes is to identify the keywords that can be used in the job advertisement to be created for the personnel need, and a solution is proposed with a word cloud. In addition, this word cloud method provides preliminary information about what those who plan to work in information technologies should know according to the field. In the second process, the resumes of the personnel sought in the job advertisement are quickly analyzed and classified with natural language processing models to ensure automatic and accurate analysis of the initial evaluations of the resumes.

The contribution of this study to the literature is as follows:

Word cloud, which is one of the natural language processing techniques, is used to determine the right keywords that can be used in recruitment advertisements.

It is a recent example of using natural language processing techniques in human resources, a fairly new field of study.

In addition to natural language processing techniques, Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM) models, this is the first study to use the Convolutional Neural Network (CNN) model, which is a deep learning model for the correct classification of CVs.

It is one of the rare studies on resume classification that shows how a model works on a resume with explainable and interpretable AI.

The next section of this paper presents the literature on resume classification. The third section presents the methodology of this study. The methodology section describes the dataset, models, and metrics used in the study. In the third section, the results obtained in this study are presented, and the results are explained in detail. In the last section, a comparison is made with the literature, and information about its future use is given.

1. Literature Review

In recent years, with the advances in artificial intelligence technologies, information technology systems have started to play a critical role in the automation of human resources processes in areas ranging from assessing higher education students to analyzing professional experience. This literature review will detail the diversity and effectiveness of methods developed for different applications, such as classification of student and job applications, curriculum vitae enhancement and error detection, knowledge extraction, and matching. The literature review presents studies on these three different applications.

Before moving on to the studies in the literature, it is necessary to mention the accuracy metric, the most frequently used metric in the studies, to make comparisons within the scope of this study. The accuracy metric refers to the number of correctly performed classification results divided by the total amount of data. The closer the accuracy value is to 100%, the better the result is considered to be.

1.1. Resume Classification and Assessment

Haddad and Mercier-Laurent (2021) conducted a study to automate the pre-selection and assessment of higher education students for use in recruitment processes. The evaluation system used classification algorithms using machine learning according to 4 main categories: personal information, academic background, professional experience, and social and technical skills. They used Naive Bayes Classifies, Support Vector Machine (SVM), and Random Forest machine learning models. The accuracy results of these models are 64%, 69%, and 78%, respectively. Pal et al. (2022) tried to classify resumes with machine learning methods. Within the scope of this study, they collected data from 3 different websites (kaggle.com, glassdoor.com and indeed.com). They used 70% of the collected data as training and 30% as testing. They used SVM, Naive Bayes, and Random Forest machine learning methods and achieved 60%, 45%, and 70% accuracy, respectively. Another study used machine learning models SVM, Naive Bayes, K-Nearest Neighbor, Logistic regression, and Term-Frequency-Inverse-Document-Frequency (TF-IDF) to classify job applications with resumes with available vacancies. For this Resume Classification System (RCS), it was observed that the SVM model used with the One-Vs-Rest-Classification strategy achieved a validation value of around 96% (Ali et al., 2022). In another study, a resume recommendation system was developed to classify and match candidate resumes with advertisements. In this study, 4 machine learning methods (Random Forest, Multinomial Naive Bayes, Logistic Regression, Linear Support Vector Machine Classifier) were used for classification. At the same time, cosine similarity and k-NN were preferred for matching. They obtained an accuracy of approximately 78% with the Linear Support Vector Machine Classifier method in resume classification (Roy et al., 2020, p. 2318).

1.2. Resume Enhancement and Error Elimination

In another study, Bert Regression, Linear Regression, Decision Tree, Support Vector Regression (SVR), and Random Forest were used to develop a Natural Language Processing (NLP) and rule-based content scoring system to eliminate errors in the resumes prepared by undergraduate students and recent graduates and to improve the resumes. Mean Absolute Error (MEE) was used to measure success, and 4 sections of the resume (Profile, Education, Projects, and Technical Skills) were analyzed. The Bert Regression model gave the best result, and the values of 0.0667 for Profile, 0.7333 for Education, 0.5333 for Project, and 1.2667 for Technical Skills were obtained (Weerasinghe et al., 2023, p.1). Bharadwaj et al. (2022, p. 238) tried to categorize CVs according to the skills they contain according to various job options using NLP and LSTM. In this study, model success was not measured with any metric. It is seen that it is provided to be analyzed with a simple interface so that every user can use it.

1.3. Information Extraction and Matching

In another study, a structure consisting of 3 parts was developed in order to extract important information from resumes and sort them according to company requirements. In the second part, they extracted meaningful data from unstructured data. In the last stage, they made an evaluation with a decision tree. Approximately 80%- 85% accuracy was obtained with the decision tree (Reza & Zaman,2022). Another study, developed for employers, ranked resumes using the word2vec algorithm, taking into account skills, experience, education, and location. Also, in this study, after ranking, resumes and employers were matched using the Gale-Shapley algorithm (Pudasaini vd.,2022). In another similar study, an application that can be updated according to the needs of an organization was developed to analyze the resumes received for a job advertisement with headings such as educational qualifications, skill sets, and technical subjects by machine learning. An application was proposed for this, but the successful results of the machine learning used were not given (Harsha et al., 2022, p. 1772).

Machine learning and natural language processing techniques were used to predict personality traits in a study conducted to analyze personal information in a CV by adding a quiz alongside the CV. The methods used were KNN, Linear Regression, Logistic Regression, SVM, and Random Forest. The best accuracy result in this study was obtained with Random Forest with 80.2% (Anusha et al., 2023, p. 1179). In another study on software engineering candidates, a character positioning technique matched words and phrases to extract candidate information. From the extracted information, a resume summary was created, and a scoring system was created based on skills. In the testing process, the results were around 33% for five random software engineering positions (Pant et al., 2022, p. 44). In another study, they used natural language processing and text mining to identify technical knowledge in software engineering positions using resumes and unstructured texts from the syllabus. They achieved an accuracy of about 98% and presented this study as a web application (Valdez-Almada et al.,2017, p.97).

2. Methodology

2.1. Dataset

The Resume-Dataset dataset published on Kaggle and created specifically for the IT sector was used (Mzali, 2023). This dataset was created by compiling job applicants' resumes from various sources, including websites like Kaggle, Glassdoor, and Indeed. This dataset consists of a CSV file. This dataset has two columns, and one contains the resume's text version. The other column contains fields related to information technologies. This dataset, which consists of more than 29000 data, consists of 6 IT-related occupational classes. These classes are database, administrator, system administrator, project manager, software developer, network administrator, and security analyst. In addition, the data set was randomly selected from the data set with the train test split method, and 80% were split for training and 20% for testing.

Before proceeding with the model training, various operations were performed on the CV dataset. These operations are as follows:

Reading the Data Set: All CVs are contained in a CSV file. As a first step, the file with CSV extension was read.

Blank Data Check: The data set was checked for empty data, and no empty data was found.

Removal of Punctuation and Special Characters: For natural language processing models to work more effectively, punctuation marks and special characters in the texts in the resume dataset were removed, leaving only letters and numbers.

Stopwords Removal Process: Removing unnecessary words (stopwords) was performed on the resume texts. Since the CVs were in English, stopwords accepted in English were removed. Examples of English stopwords; “i”, ‘me’, ‘our’, ‘your’, ‘he’, ‘him’, ‘it’, ‘them’, ‘they’, ‘what’, ‘which’, ‘who’, ‘am’, ‘is’, ‘are’, ‘be’, ‘been’, ‘being’, ‘have’, ‘has’, “had”, ‘a’, ‘an’, ‘until’, ‘of’, ‘about’, ‘with’, ‘or’, ‘and’, ‘to’, ‘from’, ‘up’, ‘all’, ‘any’, ‘both’, ‘too’, ‘can’, ‘very’, ‘should’ etc.

Converting Tags Into Numeric Values (Encoding): As a final step, the resume tags are converted so that each category tag corresponds to a numeric value. This process is called “encode”.

This was done in order to process the dataset more efficiently with natural language processing models.

2.2. Models

In this study, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) models were preferred. CNN was chosen because it is a classical deep learning model, while LSTM was chosen because it is one of the most widely used models in natural language processing applications. The GRU model is included in this study because it is an improved version of the LSTM model.

This study was performed on Google Colab Pro and developed using the Nvidia Tesla T4 GPU. The specifications of the Nvidia T4 GPU are as follows:

TFLOPS: 8.1

CUDA Cores: 2560

Memory: 16 GB GDDR6

Thanks to this powerful GPU, the training and testing of deep learning models could be done faster and more efficiently. This GPU met the project's requirements, improving the model's performance and accuracy.

2.2.1. Convolutional Neural Networks(CNN)

CNN, one of the deep learning methods, stands for "Convolutional Neural Networks". This artificial intelligence model, which is used in many fields, such as image processing, video analysis, and natural language processing, has the ability to extract features from data automatically. CNNs are particularly successful in image recognition and classification tasks. The CNN architecture consists of successive layers. These layers are;

Convolutional Layer: Performs convolutional processing on the input image to extract local features. Each convolutional operation scans different parts of the image using small filters. It then creates feature maps.

Pooling Layer: Used to reduce image size and preserve the most important information in feature maps. There are two most common pooling techniques. The first is max pooling, and the other is average pooling.

Fully Connected Layer: The features obtained after convolution and pooling are fed to one or more fully connected layers for classification. This layer learns the relationships between features and classifies objects.

Activation Function: This layer helps the model to learn non-linear features. Usually ReLU (Corrected Linear Unit) is used, but there are other activation functions such as Sigmoid, tanh, etc.

The hyperparameters of the CNN model used in this study are given in Table 1.

Table 1. Hyperparameters used in the CNN model

Hyperparameter Name	Value
vocab_size	1000
embedding_dim	50
num_filters	100
filter_sizes	[2, 3, 4]
dropout_prob	0.7
hidden_dim	64
output_size	7
lr (learning Late)	0.001
dropout	0.5
batch_size	16
num_epochs	100

Source: Authors.

2.2.2. Long Short-Term Memory (LSTM)

LSTM, a type of Recurrent Neural Network (RNN), stands for long short-term memory. Proposed in 1997 by Sepp Hochreiter and Jürgen Schmidhuber, LSTMs are specifically designed for use in sequential data processing tasks such as time series, NLP, and speech recognition. Unlike basic RNN structures, LSTM has the ability to learn long-term dependencies, making it much more suitable for complex sequential tasks.

The main feature of the LSTM is that it has a special structure that can store information for long periods of time and forget redundant information. LSTM contains 4 gates. These gates are;

Input Gate: Selects which information from the current input is important and which inputs should be added to the cell state.

Forget Gate: Decides which information to delete from the cell state. It decides the information to be deleted according to the past state and current input.

Cell State: It is seen as the heart of the LSTM. It stores long-term information. The forget gate and the enter gate can update it.

Output Gate: This gate determines the output information to be passed to the next layer or unit.

LSTM models can precisely control the flow of information through these gates. Through this control, it allows the model to learn both short-term and long-term dependencies.

The hyperparameters of the LSTM model used in this study are given in Table 2.

Table 2. Hyperparameters used in the LSTM model

Hyperparameter Name	Value
vocab_size	1000
embedding_dim	50
hidden_dim	64
output_size	7
bidirectional	True
lr (learning Late)	0.001
dropout	0.5
batch_size	16
num_epochs	100

Source: Authors.

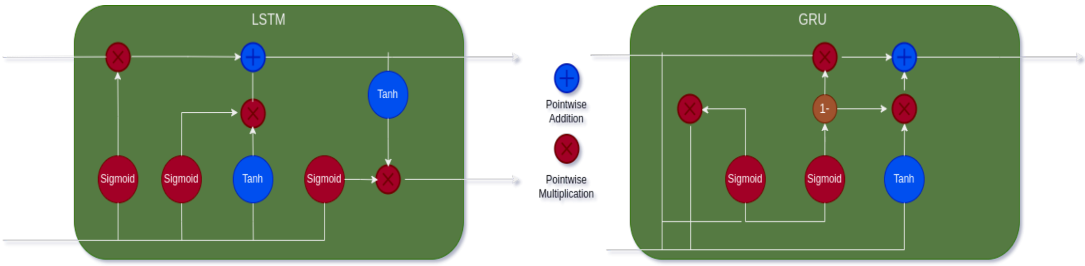
2.2.3. Gated Recurrent Unit (GRU)

Proposed in 2014 by Cho et al. GRU is a member of recurrent neural networks (RNN). Developed to overcome the difficulties of RNNs in learning long-term dependencies, GRU achieves high success in natural language processing and time series data.

Working in a similar way to the LSTM model, the main goal of GRU is to provide a structure that can remember information for long periods of time and forget redundant information. However, accomplishing this task offers a simpler structure than LSTM. Unlike the LSTM, it has 2 gates. One of the gates is the Reset Gate. This gate decides how much of the past information will be used in the creation of the future state and how this information will be combined. The other gate is known as the Update Gate, which determines which information to keep or discard. It sets the degree of importance between past information and new inputs and decides which information to pass on to the next state.

GRU works faster and more efficiently than LSTM. The biggest reason for this is that it has fewer parameters than LSTM. It is preferred when learning resources are limited (Figure 1).

Figure 1. Diagram of LSTM and GRU models



Source: Authors.

The hyperparameters of the GRU model used in this study are given in Table 3.

Table 3. Hyperparameters used in the GRU model

Hyperparameter Name	Value
vocab_size	1000
embedding_dim	50
hidden_dim	64
output_size	7
lr (learning Late)	0.001
batch_size	16
num_epochs	100

Source: Authors.

2.2.4. Local Interpretable Model-agnostic Explanations (LIME)

LIME is a method for making the decisions of complex machine learning models more understandable. LIME is considered a tool for both explainable AI and interpretable AI. explainable AI is a technique that provides transparency and ease of understanding in the decisions made by AI models. Interpretive AI is another technique for making the internal logic and decision-making process of an AI model understandable. The main features of LIME are;

Model independent. It can be applied to every model and can show the working logic of each model.

Understandability: It helps to understand the predictions of complex models.

Local Approach: LIME examines a small region around a single forecast.

The LIME approach works in 3 basic steps.

Firstly, it creates many synthetic data points in the data space around the original data points.

The complex model then predicts the synthetic data points.

Based on the predictions of the complex model, LIME creates a local description with a simple and straightforward model.

In this study, the models we used to classify resumes with LIME are used in the prediction process to see how a resume is analyzed. The class-specific weights of the words in the resume are also included.

2.3. Metrics

In this study, the success of the models in resume classification is measured by 4 metrics. The first of these metrics is the accuracy metric. The accuracy metric is one of the most commonly used performance measures in machine learning and statistics, particularly for classification problems. It measures how well a model's predictions match the true values. Accuracy is calculated as the total number of correctly predicted examples (both positive and negative) divided by the total number of examples. The accuracy metric can be misleading in the presence of unbalanced classes. Summarizes the performance of a model in a simple way (Hossin et al., 2015, p. 1). The closer the accuracy value is to 100, the better the result.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

In this equation,

True Positive = Refers to the situation where the forecast and actual value is YES.

True Negative = Refers to the situation where the forecast and actual value is NO.

False Positive = It refers to the situation where the model prediction is YES, but the actual output is NO.

False Negative = It refers to the situation where the model prediction is NO, but the actual output is YES.

Another metric, Mean Squared Error (MSE), is an important metric used in statistics and machine learning to evaluate model performance. It measures the amount of error by averaging the square of the difference between the predicted values and the true values. Low MSE values indicate that the model makes predictions close to the true values, while high MSE values indicate poor model performance (Wallach & Goffinet, 1989, p. 299). The ideal value for this metric is 0. The closer it is to 0, the more perfect the classification. However, the ideal value of the MSE metric may vary depending on the classification problem. For example, in a house price prediction application, a margin of error of 1000 Turkish liras among houses worth millions of Turkish liras can be considered ideal. Therefore, the ideal value for MSE is related to the problem itself.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i) \quad (2)$$

In this equation,

n is the number of samples,

Y_i actual values,

$\hat{Y}_i$ represents the values predicted by the model.

MCC (Matthews Correlation Coefficient) is a metric commonly used to evaluate the performance of a model in classification problems. MCC expresses the performance of a classification model with a value between -1 and 1, taking into account imbalances between all classes of a classification model. The expected value of a model making completely random predictions is 0, while it takes the value 1 or -1 for perfect classification. -1 denotes a complete misclassification (Baldi et al., 2000, p.413). The main advantage of the MCC metric is that it accurately measures model performance even with imbalanced datasets. The best value for MCC is close to +1. A value of +1 indicates correct classification. A value of 0 indicates that the model produces a random result. A value of -1 indicates that the model completely misclassifies.

$$MCC = \frac{(TP\ x\ TN\ )-(FP\ x\ FN)}{\sqrt{(TP+FP)x(TP+FN)x(TN+FP)x(TN+FN)}} \quad (3)$$

Another metric used in modeling imbalanced datasets is the F_1 Score, a metric used to evaluate the performance of the model in classification tasks. The F_1 Score is calculated as the harmonic mean of precision and recall and takes a value between 0 and 1. Precision is the probability that the instances that the model predicts as positive are actually positive. At the same time, recall indicates how many of the instances that are actually positive are correctly predicted as positive (Goutte & Gaussier, 2005, p. 345). The ideal value for precision, recall, and F_1 Score is 1. Precision 1 means no false positives, while recall 1 means no false negatives.

If the F_1 Score is 1, it means that there is perfect agreement between precision and recall.

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$F_1\ Score = 2\ x \frac{Precision\ x\ Recall}{Precision+Recall} \tag{6}$$

3. Results

In the first phase of this study, we used the word cloud method, a Natural Language Processing (NLP) method, to identify required qualifications for a field from resumes in the field of computer technology. A word cloud is a visual representation of the words in a collection of text or documents. Frequent words in the text are shown in larger and/or bolder fonts, while less frequent words are shown in smaller fonts. With this method, general information about a document can be obtained. In this study, the word cloud method was used to obtain information about the fields of information technologies by analyzing resumes in information technologies. The common words in the documents belonging to the same class (information technology fields) have been identified, keywords that can be used in new job postings have been identified, and it provides preliminary information about what those who want to work in that class should know.

Figure 1 shows the word clouds created for 6 information technologies. In the word cloud created for the Database administrator role in Figure 1, it is seen that concepts such as Oracle Database, SQL, high availability, Server, and backup recovery are prominent. However, for the same role, it is seen that concepts such as 10g, customer service, and education are mentioned very little. When analyzed for the System Administrator role, it is seen that concepts such as system administrator, active directory, security, and windows server are mentioned very often, while concepts such as employee, multiple, group policy, and virtual machine are mentioned less frequently. For another role, project manager, concepts such as work us, financial, and assisted are used less frequently, while concepts such as support, project management, work experience, report, managed, application, client, and lead are mentioned quite frequently. When the software developer role is analyzed, concepts such as implemented, software engineer, visual studio, SQL server, data, and web service come to the fore. For the same role, application use, enterprise, standard, group, organization, and issue are mentioned less frequently. For network administrators, which is a sub-role of information technologies, concepts such as active directory, customer service, configuration, network engineering, and technical support are frequently mentioned, while concepts such as deployment, installed, supported, maintain are rarely mentioned. When Security analyst, which is the last role in this study, is examined, it is seen that security-oriented concepts such as information security, risk assessment, system, client, active directory, security control, and cyber security stand out. In contrast, concepts such as help desk, hardware, and configuration are rarely used.

When the word clouds given in Figure 2 are analyzed, it is seen that some concepts belonging to some roles are common, while some concepts are specific to that role. For example, while the concept of risk assessment and cyber security is specific to the security analyst, the concept of active directory is a common concept with the system administrator. Likewise, concepts such as Oracle Database and backup recovery are specific to the database administrator, while SQL server is common with the software developer role. The concept of application is common to the professions of project manager and software developer, while the concept of web service is specific to the software engineer profession. The results of the models used in this study are given in Table 4. When Table 4 is analyzed, it is seen that the best result in all metrics is obtained in the GRU model. An accuracy rate of just over 99% is a very good result. In addition, the GRU model obtained 0.9883 in the MCC metric, 0.9865 in F_1 score and 0.1119 in MSE. The best value for accuracy, MCC and F_1 score metrics is 1, while the best value for MSE is 0. The GRU model was closest to 0 in MSE and closest to 1 in other metrics. The closest accuracy value to the GRU model was CNN with 97.95%. CNN model gave values of 0.9743 in MCC metric, 0.9794 in F_1 score metric and 0.1278 in MSE metric. The worst result was the LSTM model with an accuracy of 92.35%. LSTM gave values of 0.4628 in MSE metric, 0.9042 in MCC and 0.9168 in F_1 score. Loss values are also given in Table 4. When all metrics are analyzed, it is seen that the GRU model gives the best result with a value of 0.03 in terms of loss values. With this value, it is clearly seen that the GRU model shows a success parallel to its success in all metrics in the loss value. Loss values of other models are 0.2819 for the LSTM model and 0.0754 for CNN.

Figure 2. Extracting Qualifications According to Their Fields from the Resumes of Candidates in the Field of Information Technologies by Word Cloud Method

Source: Authors.

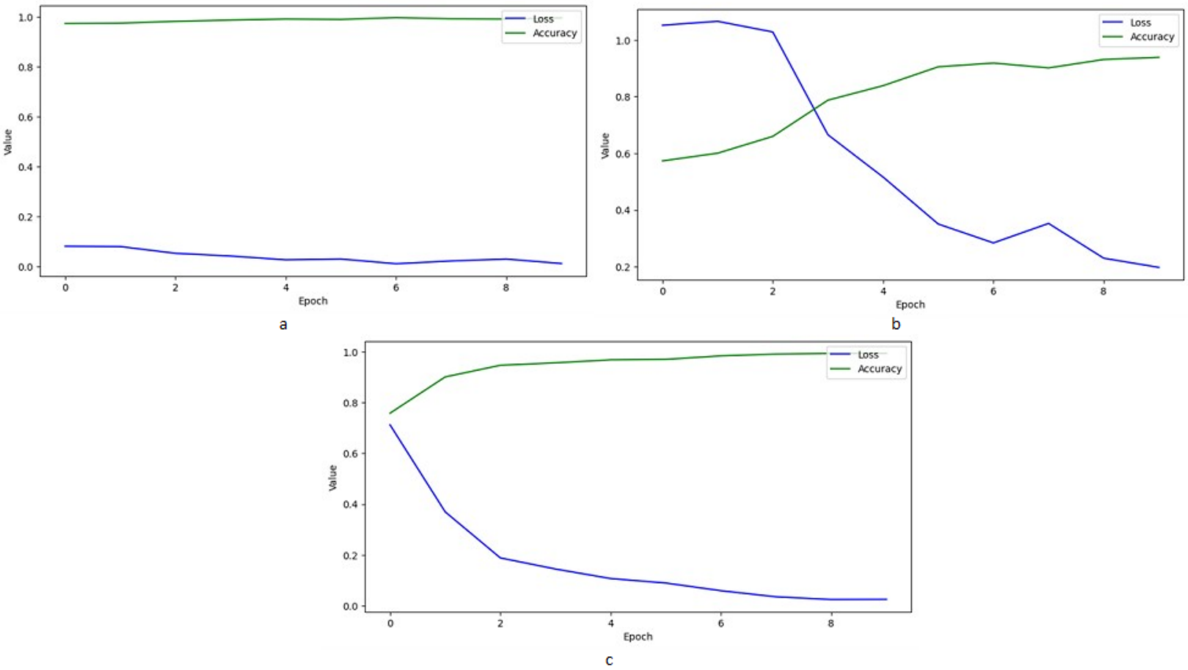
In addition, Figure 3, attached to Table 4, shows the changes in the accuracy and loss metrics of the models during the training process. At the end of the training process, the loss value is expected to be low, and the accuracy value is expected to be high. During the training process, the loss value gradually decreases, and the accuracy value increases, indicating that the training is healthy. It is seen that there is no significant change in the accuracy and loss values of the CNN model during the training process. In other words, the CNN model has been quite stable throughout the training process. When the graph of the LSTM model is analyzed, it is seen that while the values were very bad at the beginning, they improved as the training process progressed. The GRU model, which gave the best result, started with worse values than the CNN model initially but improved as the training process progressed. During the training process, LSTM was the most improved model, while CNN was the least improved model.

Table 4. Results of Models

Metrics					
Models	**Accuracy (%)**	**MCC**	**F_1 Score**	**MSE**	**Loss**
CNN	97.95	0.9743	0.9794	0.1278	0.0754
LSTM	92.35	0.9042	0.9168	0.4628	0.2819
GRU	99.07	0.9883	0.9865	0.1119	0.03

Source: Authors.

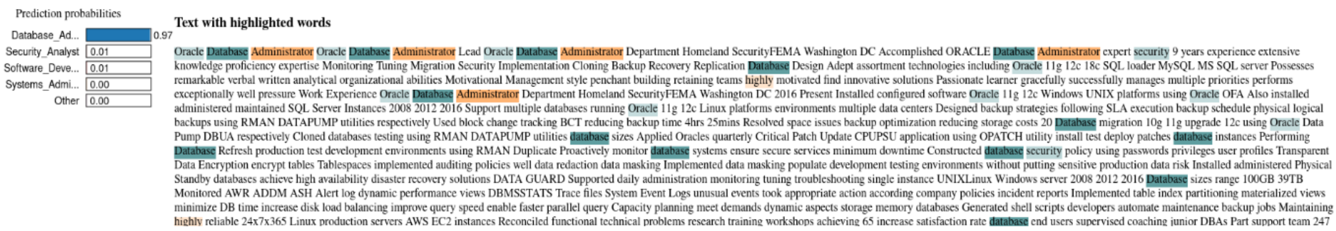
Figure 3. Plot of Accuracy and Loss Results For 9 Epochs *(a: CNN, b: LSTM, c: GRU)*



Source: Authors.

LIME was used to analyze a resume of the Database_Administrator class of the GRU model. The result of this analysis is shown in Figure 4. Figure 4 should be analyzed in two stages. First, the right side of Figure 4 shows the text information of the resume. When this text information is examined, some words on the text are highlighted in color. These highlights show which words the model finds important and which words are taken into account in the prediction process and are effective in predictions. For example; the word “Database” is highlighted in green and repeated quite frequently in the text. In addition, the word “Administrator” is highlighted with orange color in the text. The word “oracle” and “security” are highlighted with a similar color tone, showing that the model is equally effective in the decision-making process. It is concluded that these 4 words (“Database”, “Administrator”, “oracle”, “security”) are effective in the decision making process. When the left side of Figure 4 is analyzed, it can be concluded to which classes the given resume can belong. 97% of the time “Database_Administrator” makes the decision, 1% value is calculated for “Security_Analyst” and “Software_Developer” class, while the value of the other classes is 0.

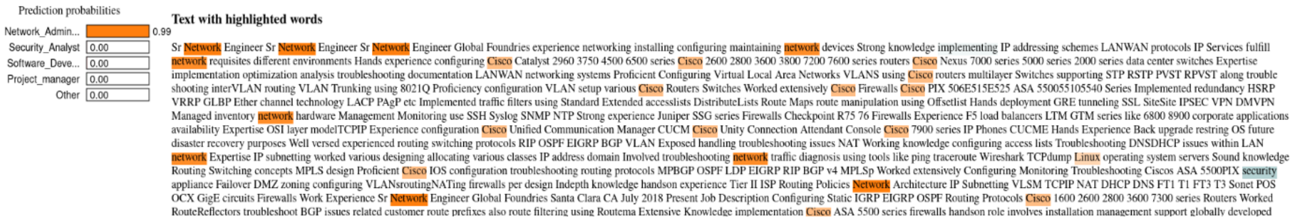
Figure 4. The Main Application of LIME and the GRU Model on the Decision-Making Process of a Resume Belonging to the Database_Administrator Class



Source: Authors.

LIME was used to analyze a resume of the Network_Administrator class of the GRU model. The result of this analysis is shown in Figure 5. Figure 5 should be analyzed in two stages. First, the right side of Figure 5 shows the text information of the resume. When this text information is examined, some words in the text are highlighted in color. These highlights show which words the model finds important and which words are taken into account in the prediction process and are effective in predictions. For example, the most effective word is “network”, which is highlighted in orange. The word “Cisco” highlighted in light orange and the word “security” highlighted in blue are also effective in the classification decision process. When the left side of Figure 5 is analyzed, it is seen that the model made the decision of “Network_Administrator” with 99% accuracy. When the other classes are analyzed, it is seen that the rate is 1% and lower.

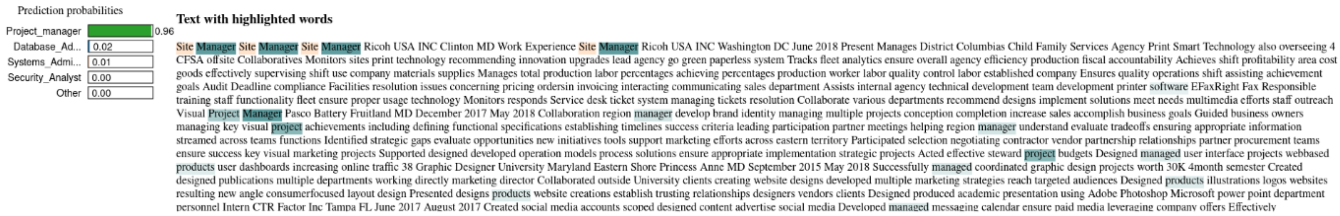
Figure 5. The Main Application of LIME and the GRU Model on the Decision-Making Process of a Resume Belonging to the Network_Administrator Class



Source: Authors.

LIME was used to analyze a resume of the "Project_manager" class of the GRU model. The result of this analysis is shown in Figure 6. Figure 6 should be analyzed in two stages. Firstly, the right side of Figure 6 shows the text information of the resume. When this text information is examined, some words in the text are highlighted in color. These highlights show which words the model finds important and which words are taken into account in the prediction process. For example, it is seen that the words "manager" and "project" highlighted in green are the most effective words in the decision making process. The words "products", "managed", "manager" and "site" highlighted in light green are also effective in the decision-making process. As a result of the analysis of these words, the GRU model predicted the "Project_manager" class 96% of the time, while it predicted the "Database_Administrator" class 2% of the time. The GRU model calculated that the other classes have 1% and less.

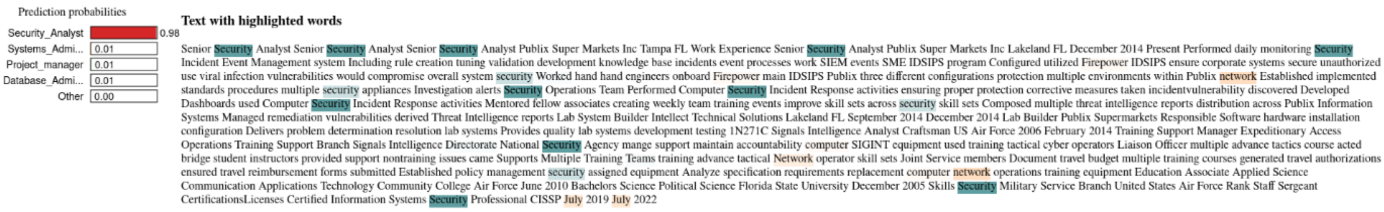
Figure 6. The Main Application of LIME and the GRU Model on the Decision-Making Process of a Resume Belonging to the Project_manager class



Source: Authors.

LIME was used to analyze a resume of the Security_analyst class of the GRU model. The result of this analysis is shown in Figure 7. Figure 7 should be analyzed in two stages. First, the right side of Figure 7 shows the text information of the resume. When this text information is examined, some words in the text are highlighted in color. These highlights show which words the model finds important and which words are taken into account in the prediction process. For example, the word "Security" is highlighted in green and it is seen that it is the most effective word. When we look at the other words, it is seen that the words "network" and "July" are also effective. When the left side of the Figure is analyzed, it is seen that the GRU model decided that 98% of the CV belongs to the "Security_analyst" class. In addition, the GRU model calculated the probability of belonging to the "System_Administrator", "Project_manager" and "Database_Administrator" classes as 1%, while the value for the other classes was 0.

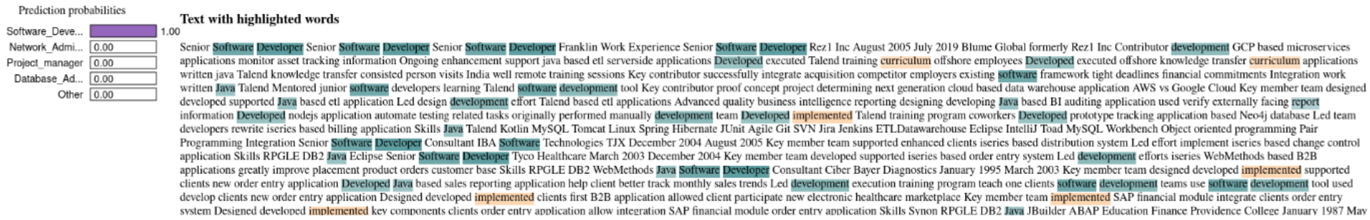
Figure 7. The Main Application of LIME and the GRU Model on the Decision-Making Process of a Resume Belonging to the Security_Analyst class



Source: Authors.

LIME was used to analyze a CV of the Software_Developer class of the GRU model. The result of this analysis is shown in Figure 8. Figure 8 should be analyzed in two stages. First, the right side of Figure 8 shows the text information of the resume. When the CV is examined, it is seen that the words "Software_Developer", "Developed", "Java" and "development" highlighted in green are the most effective words. The words "curriculum" and "implemented" highlighted in orange are effective for this class but may be effective for other classes. The GRU model has determined with 100% certainty that this resume belongs to the "Software_Developer" class.

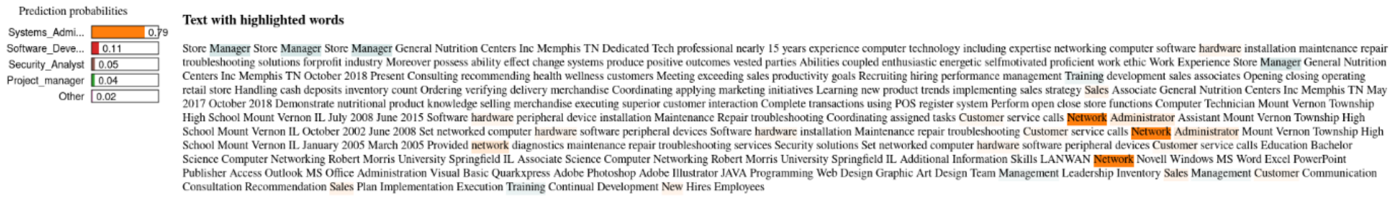
Figure 8. The Main Application of LIME and the GRU Model on the Decision-Making Process of a Resume Belonging to the Software_Developer class.



Source: Authors.

LIME was used to analyze a resume of the System_Administrator class of the GRU model. The result of this analysis is shown in Figure 9. Figure 9 should be analyzed in two stages. First, the right side of Figure 9 shows the textual information of the resume. When the CV is examined, it is seen that only one word is expressed in orange color and this word may be effective for other classes. The word shown in orange is “network”. In this case, it is not possible to talk about the effect of one word clearly, but it is seen that there are very few effective words. These words are words like “manager”, “administrator”, “sales”. GRU concluded that this resume belongs to the System_Administrator class with a slightly lower value (79%), but with 11% it belongs to the Software_Developer class. Regarding the other classes, the results show that 5% belong to “Security_Analyst”, 4% to “Project Manager” and 2% to other classes.

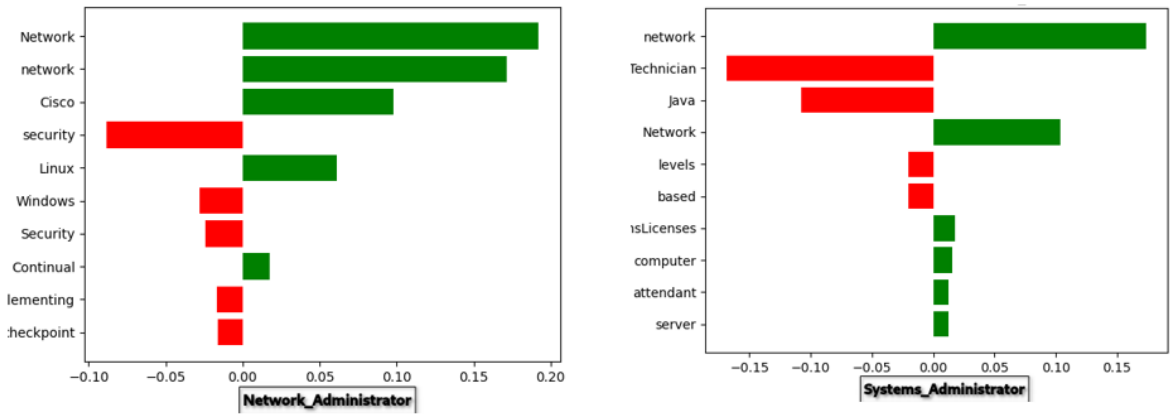
Figure 9. The Main Application of LIME and the GRU Model on the Decision-Making Process of a Resume Belonging to the System_Administrator class.



Source: Authors.

Figure 10 shows the words used by the GRU model in the classification process for Network_Administrator and System_Administrator classes with LIME. When examining the bar graphs in Figure 10, it should be noted that green colors indicate that the words belonging to those classes are important words, while red colored words have a negative impact for the class. In Figure 10 (a), it can be seen that the words “Network”, “network”, “Cisco”, “Linux” and “Continual” belonging to the Network_Administrator class are effective in the decision making process. However, the words “security”, “Windows”, “Security”, “Security”, “implementing” and “checkpoint” belonging to a different class may have an effect on the conclusion, while it is negative for the Network_Administrator class. In Figure 10 (b), the important words for the Systems_Administrator class are “network”, “Network”, “Licenses”, “computer”, “attendant” and “server”. The negative words belonging to this class are “Technician”, “Java”, “levels” and “based”. Among the important words in these two classes, the words “network” and “network” are common. In addition to these common words, the GRU model also looks at other important words to determine which class they belong to. So, if the word “Linux” or “Cisco” appears next to the word “Network” in the resume, the GRU model decides that this resume belongs to the Network_Administrator class. On the other hand, if the words “computer” or “server” appear next to “network” or “Network”, the GRU model decides that this resume belongs to the System_Administrator class.

Figure 10. Word Analysis with LIME for Network_Administrator and System_Administrator Classes of GRU model



(a) (b)

Source: Authors.

Conclusion

The word cloud techniques used in this study have been shown to reveal the qualities that candidates who will work in computer technology should know and have according to the roles they want. This will help the candidate develop himself/herself according to the role beforehand and give the chance to organize his/her training according to these characteristics. In addition, word clouds can be used as an auxiliary application that can be used by the human resources department in the recruitment process, from the job advertisement to the end of the pre-assessment process, providing preliminary information about which qualifications they should have according to the role. The effective use of the word cloud method can provide significant contributions both for human resources and candidates.

In this study, the GRU model used in the resume classification step achieved an accuracy of just over 99%. In other words, the GRU model performs a very accurate classification. In addition, when Table 4 is examined, the closest study to the GRU model in the classification of general resumes is the study by Haddad and Mercier-Laurent (2021), where they obtained 85% accuracy with the Decision Tree model. On IT resumes, the closest result to this study is the study by Valdez-Almada et al. (2017, p.97), where they obtained a 98% accuracy value with the text mining method. There may be two major reasons why our study yielded the best results. The first reason may be that we only worked with resumes related to information technologies. The second reason is that we used current and newly developed models. We believe that our results are better than those of other studies because we work with current models, such as GRU, among natural language processing models.

The research results show that these deep learning models provide high accuracy rates and efficiency in human resources' resume classification and candidate matching processes. Furthermore, the paper shows that applying the models can significantly help human resource professionals in candidate selection by improving the speed and quality of recruitment processes. These models can be integrated into applications such as CRM, making human resources modules AI-enabled.

Table 4. Comparison with the Literature

Literature	Model	Accuracy (%)	Dataset Feature
Haddad and Mercier-Laurent (2021)	Random Forest	78	General
Reza and Zaman(2022)	Decision Tree	85	General
Anusha et al. (2023, p. 1179)	Random Forest	80.2	General
Roy et al. (2020, p. 2318)	Linear Support Vector Machine Classifier	78	General
Pant et al. (2022, p. 44)	Extract Information	33	IT
Valdez-Almada et al. (2017, p. 97)	Text Mining	98	IT
Pal et al. (2022)	Random Forest	70	General
Our Proposal	***GRU***	***99***	***IT***

Source: Authors.

Finally, this study has shown that the word cloud method can be used to select the best keywords for advertisements in information technology. This method helps candidates to make decisions during the application process and shows what students studying IT should know according to the field they want to work in. In addition, the curricula of departments such as computer engineering and software engineering can be reorganized according to these keywords, and the personnel needed by the sector can be trained.

There are some limitations in this study. The data set used in this study includes only CVs in information technologies. There are limitations regarding data set diversity and number of data. The GRU model, which gives the best results as a model, gave good results on IT resumes. However, how it will perform in general resume classification or in different data sets remains to be seen. In addition, the study did not test integration and compatibility with the tools used in human resources processes.

The study can be repeated in future studies with more extensive and diverse datasets, including different sectors and job roles. In addition, the universality of the study can be extended by training with resumes in different languages, not only English resumes. In addition to data set studies, modifications can be made to the models, or improvements can be made to new models. This AI-based study and similar studies can be integrated with human resource management systems (HRMS) and customer relationship management systems (CRM) to test the models in real environments. In addition, model results can be verified by a human resources expert. The accuracy can be increased by repeating the training process with user feedback.

Author Contributions
The authors have declared contributions to research as follows: Selim Sürücü: 50%, Berk Küçük: 25%, Mustafa Kemal Aydın: 25%

Acknowledgments
The authors reported no acknowledgment.

Funding and Support
The authors reported no funding and support.

Conflict of Interests
The authors reported no conflict of interest.

Ethics Statement
The authors have reported no need for ethical committee approval.

Corresponding Editors
Assoc. Prof. Dr. Beyza Mina Ordu-Akkaya, Ankara University of Social Sciences
Assoc. Prof. Dr. Görkem Ataman, Yaşar University
Res. Asst. Yunus Yıldırım, Afyon Kocatepe University

References

Ali, I., Mughal, N., Khand, Z. K., Ahmed, J., & Mujtaba, G. (2022). Resume classification system using natural language processing and machine learning techniques. *Mehran University Research Journal Of Engineering & Technology, 41*(1), 65-79.

Anusha, K. O. V., Dhar, A., Dixit, S., Saraf, A., & Lonial, I. A. N. (2023). Automated personality-based candidate shortlisting using machine learning and natural language processing. *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, Trichy, India, 2023, pp. 1179-1184. https://doi.org/10.1109/ICAISS58487.2023.10250633

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A. F., & Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: An overview. *Bioinformatics, 16*(5), 412-424.

Bharadwaj, S., Varun, R., Aditya, P. S., Nikhil, M., & Babu, G. C. (2022). Resume screening using NLP and LSTM. *2022 International Conference on Inventive Computation Technologies (ICICT)*, Nepal, 2022, pp. 238-241. https://doi.org/10.1109/ICICT54344.2022.9850889

Cho, K., Merriënboer, B. V., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724-1734. https://doi.org/10.3115/v1/D14-1179

Haddad, R., & Mercier-Laurent, E. (2021). Curriculum vitae (CVs) evaluation using machine learning approach. In E. Mercier-Laurent, M. Ö. Kayalica, & M. L. Owoc (Eds.), *Artificial intelligence for knowledge management* (Vol. 614). Springer, Cham. https://doi.org/10.1007/978-3-030-80847-1_4

Harsha, T. M., Moukthika, G. S., Sai, D. S., Pravallika, M. N. R., Anamalamudi, S., & Enduri, M. (2022). Automated resume screener using natural language processing (NLP). *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, 2022, pp. 1772-1777. https://doi.org/10.1109/ICOEI53556.2022.9777194

Hossin, H., & M. N., S. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process, 5*, 01–11. https://doi.org/10.5121/ijdkp.2015.5201

Goutte, C., & Gaussier, E. (2005). A probabilistic interpretation of precision, recall, and F-score, with implication for evaluation. *Advances in Information Retrieval*, pp. 345-359.

Mezhoudi, N., Aşghamdi, R., Aljunaid, R., Krichna, G., & Düştegör, D. (2021). Employability prediction: A survey of current approaches, research challenges and applications. *Journal of Ambient Intelligence and Humanized Computing, 14*, 1489-1505. https://doi.org/10.1007/s12652-021-03276-9

Mzali, W. (2023, February 27). Resume dataset. Retrieved on 18 December 2023 from https://www.kaggle.com/datasets/wahib04/multilabel-resume-dataset

Pal, R., Shaikh, S., Satpute, S., & Bhagwat, S. (2022). Resume classification using various machine learning algorithms. *In ITM Web Conf.* https://doi.org/10.1051/itmconf/20224403011

Pant, D., Pokhrel, D., & Poudyal, R. (2022). Automatic software engineering position resume screening using natural language processing, word matching, character positioning, and regex. *2022 5th International Conference on Advanced Systems and Emergent Technologies (ICASET)*, Hammamet, Tunisia, 2022, pp. 44-48. https://doi.org/10.1109/IC_ASET53395.2022.9765916

Pudasaini, S., Shakya, S., Lamichhane, S., Adhikari, S., Tamang, A., & Adhikari, S. (2022). Scoring of resume and job description using Word2vec and matching them using Gale–Shapley algorithm. In I. Jeena Jacob, F. M. Gonzalez-Longatt, S. Kolandapalayam Shanmugam, & I. Izonin (Eds.), *Expert clouds and applications* (Vol. 209). Springer, Singapore. https://doi.org/10.1007/978-981-16-2126-0_55

Reza, Md. T., & Zaman, Md. S. (2022). Analyzing CV/resume using natural language processing and machine learning.

Roy, P. K., Chowdhary, S. S., & Bhatia, R. (2020). A machine learning approach for automation of resume recommendation system. *Procedia Computer Science, 167*, 2318-2327. https://doi.org/10.1016/j.procs.2020.03.284

Smith, J. D. (2021). Recruitment strategy. In A. B. Johnson & C. D. Lee (Eds.), *The Oxford Handbook of Human Resource Management* (pp. 123-145). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198833783.001.0001

Poulose, S., Bhattacharjee, B., & Chakravorty, A. (2024). Determinants and drivers of change for digital transformation and digitalization in human resource management: A systematic literature review and conceptual framework building. *Management Review Quarterly*. https://doi.org/10.1007/s11301-024-00423-2

Valdez-Almada, R., Rodriguez-Elias, O. M., Rose-Gomez, C. E., Velazquez-Mendoza, M. D. J., & Gonzalez-Lopez, S. (2017). Natural language processing and text mining to identify knowledge profiles for software engineering positions: Generating knowledge profiles from resumes. *2017 5th International Conference in Software Engineering Research and Innovation (CONISOFT)*, Merida, Mexico, 2017, pp. 97-106. https://doi.org/10.1109/CONISOFT.2017.00019

Wallach, D., & Goffinet, B. (1989). Mean squared error of prediction as a criterion for evaluating and comparing system models. *Ecological Modelling, 44*(3-4), 299-306. https://doi.org/10.1016/0304-3800(89)90035-5

Weerasinghe, R. L., Perera, N. N., Warusawithana, S. P., Hindakaraldeniya, T. M., & Ganegoda, G. U. (2023). 2023 8th International Conference on Information Technology Research (ICITR). *2023 8th International Conference on Information Technology Research (ICITR)*, Colombo, Sri Lanka, 2023, pp. 1-6. https://doi.org/10.1109/ICITR61062.2023.10382755