



كلية الحاسبات والذكاء الاصطناعي  
faculty of computer and Artificial Intelligence  
كلية معتمدة

## **An Arabic Model for Recruitment Fraud Detection**

A thesis submitted to the Faculty of Computers and Information - Fayoum University in fulfillment of the requirements for the M. Sc. degree in Information Systems

**Submitted by:**

**Mohamed Ahmed Sofy Mohamed**

Information System Department

Faculty of Computer and Artificial Intelligence, Fayoum University

**Supervised by:**

**Prof. Mohamed Helmy Khafagy**

Computer Science Department,  
Dean of the Faculty of Computer and Artificial  
Intelligence, Fayoum University

**Dr. Rasha Mohamed Badry**

Information Systems Department,  
Faculty of Computer and Artificial Intelligence,  
Fayoum University

2022

## **Abstract**

Over the last years, with the tremendous growth of digital transformation and the constant need for companies to hire employees, huge amounts of fraudulent jobs have been posted on the internet. A cleverly planned sort of scam aimed at job searchers for a variety of unprofessional purposes is a false job posting. It can lead to a loss of money and effort. An Arabic intelligent model has been built to avoid fraudulent jobs on the Internet using machine learning, data mining, and classification techniques. The proposed model is applied to the Arabic version of The Employment Scam Aegean (EMSCA) dataset. It is available on the Internet in English version and it has been retrieved from the use of a real-life system and consists of several features such as company profile, company logo, interview questions, and more features depending on job offer ads.

Four experiments have been conducted using the translated EMSCA dataset. Full-featured experiment, that works on all features in the dataset. Top five weighted featured experiment, applied weight for each feature to study what feature has a significant impact. Benefits-processing feature experiment, applied text processing for features, and top five weighted features with benefits extraction Experiment, that combine between second and third experiment working on top feature weight and positive keywords extraction from benefits feature. Using a set of different classifiers, such as Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes (NB), and K-Nearest Neighbor (KNN), was used to detect the fraudulent jobs. Finally, the results were compared to determine the best classifier used for detecting fraudulent jobs. The proposed model achieved better results in the benefits-processing feature experiment using a random forest classifier with 97% accuracy.