Title - A machine learning approach to detecting fraudulent job types

Publish Year- 2022

selection of data- EMSCAD

Data annotation - type (real, identity, corporate, marketing, corporate identity)

feature selection- 1) bag of words models

2) empirical ruleset

3) word embedding

4) transformers

Algorithms - LR,SGD,KNN,SVM,RF,AB,CART

Title - Online Recruitment Fraud Detection: A Study on Contextual Features in Australian Job Industries.

Publish Year – 2022

Dataset – EMSCAD, Gumtree(2730 jobs) ,

Fraud jobs - 495/2730 (18 %) .

Binary feature vector generation and generation of contextual features.

Multi-layer detection framework.

* Data collection
* Data analysis and fraud detection
* Contextual feature extraction

Experiment – Content based features, content based + contextual features

Algorithms – RF, NB, DT, rule-based classifier

Title - Effective Identification of Spam Jobs Postings Using Employer Defined Linguistic Feature

Publish Year – 2022

Dataset – Kaggle – Fake job posting prediction

* There are total 17 independent and 1 dependent features in dataset but proposed system uses only 6 features.
* Title
* Department Information
* Company Description
* Job description
* Jon Requirement
* Benefits

Algorithms – RF, SVM, Bi-directional(LSTM)

Title - Web Scraping Techniques and Applications: A Literature Review

Publish year – 2021

Methods:

* Traditional copy and paste
* HTML parsing
* DOM parsing
* Regular Expression
* XPath
* Vertical aggregation platform
* Semantic annotation recognizing
* Computer vision web page analyzer

Web scrapping Technology

Web crawler :

* Hidden
* Distributed
* Incremental
* Parallel
* Focused

Web scrapping parsers:

Web scrapping tools:

* Using PHP
* Using Python libraries
* Using java libraries
* Using Selenium
* Using apache Nutch