MAJOR PROJECT
ON
FAKE JOB POST PREDICTION USING ML & DL

BY

R.POOJITHA(19WJ5A0527)

Under the Guidance of

Mrs. CH. SUSHMA

Asst.professor,Cse Dept

CONTENTS

ABSTRACT

- This project proposed to use random forest classifier and deep neural network to predict a job post if it is real or fraudulent.
- We have experimented on Employment Scam Aegean Dataset (EMSCAD) containing 18000 samples. Deep neural network and Random Forest Classifier as classifiers, performs great for this classification task.
- We have used three dense layers for this deep neural network classifier. The trained classifier shows approximately 98% classification accuracy (DNN) to predict a fraudulent job post.

EXTISTING SYSTEM

- Social media and advertisements in electronic media have created newer and newer opportunity to share job details.
- So, to reduce the fake job posts, here they used Text Processing to differentiate true and fake posts & Dataset Training.
- Zhang proposed an automatic fake detector model to distinguish between true and fake news (including articles, creators, subjects) using text processing. They had used a custom dataset of news or articles.
- This dataset was used to train the proposed GDU diffusive unit model. Receiving input from multiple sources simultaneously, this trained model performed well as an automatic fake detector model.

DRAWBACKS OF EXISTING SYSTEM

- Less Classification Accuracy
- Less Accuracy
- Less Precision Rate

PROPOSED SYSTEM

- This project proposed to use random forest classifier and deep neural network to predict a job post if it is real or fraudulent.
- We have experimented on Employment Scam Aegean Dataset (EMSCAD) containing 18000 samples which contains real life fake job posts
- ML classifiers where our work dataset is trained. DNN creates a number of virtual neurons initialized with a random numerical value as connection weights.
- This weight is multiplied with the input and produce an output between 0 and 1. The training process adjust the weights to classify the output efficiently.

PROPOSED SYSTEM ADVANTAGES

- More Classification Accuracy
- More Accuracy
- More Precision Rate

SYSTEM ARCHITECTURE

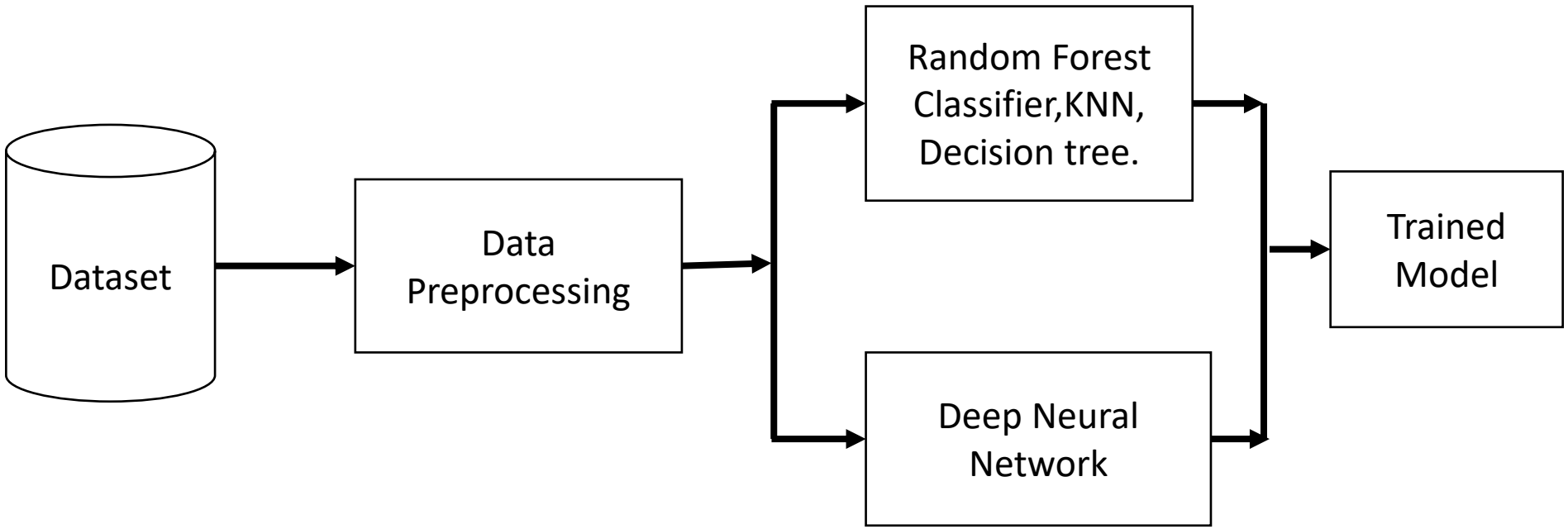


Fig: System Architecture

MODULES NAMES:

- **Data Collection**
- **Dataset**
- **Data Preparation**
- **Model Selection**
- **Analyze and Prediction**
- **Accuracy on test set**
- **Saving the Trained Model**

MODULES

- **Data Collection:**

- This is the first real step towards the real development of a machine learning model, collecting data.

- There are several techniques to collect the data, like web scraping, manual interventions and etc.

- **Dataset**

- The dataset consists of 17880 individual data. There are columns in the dataset,
- which are described below job id, title,location, department, salary range, profile, description, requirements, benefits, telecommuting,has company logo etc,

MODULES

- **Data Preparation**
- Wrangle data and prepare it for training.
- Clean that which may require it (remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.)
- Visualize data to help detect relevant relationships between variables
- Split into training and evaluation sets

- **Model Selection:**

- We used Random Forest Classifier algorithm,
- We got a accuracy of 95.02% on test set so we implemented this algorithm

MODULES

- **Analyze and Prediction:**
    - In the actual dataset, we chose only 2 features:

1. Description - detailed description of the job advertisement
2. Fraudulent - indicates whether the job is fraudulent

- **Accuracy on test set:**
    - We got an accuracy of 95.02% on test set.

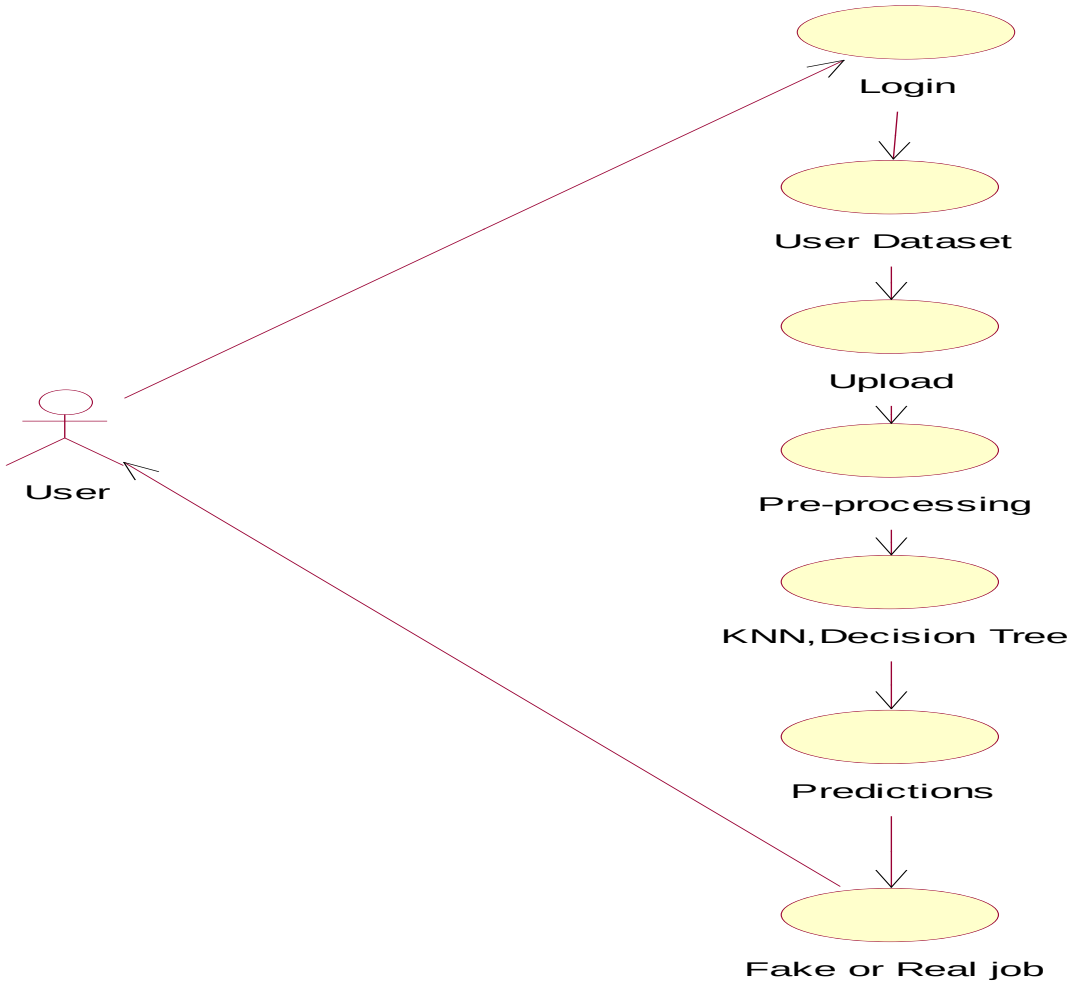
MODULES

- **Saving the Trained Model:**

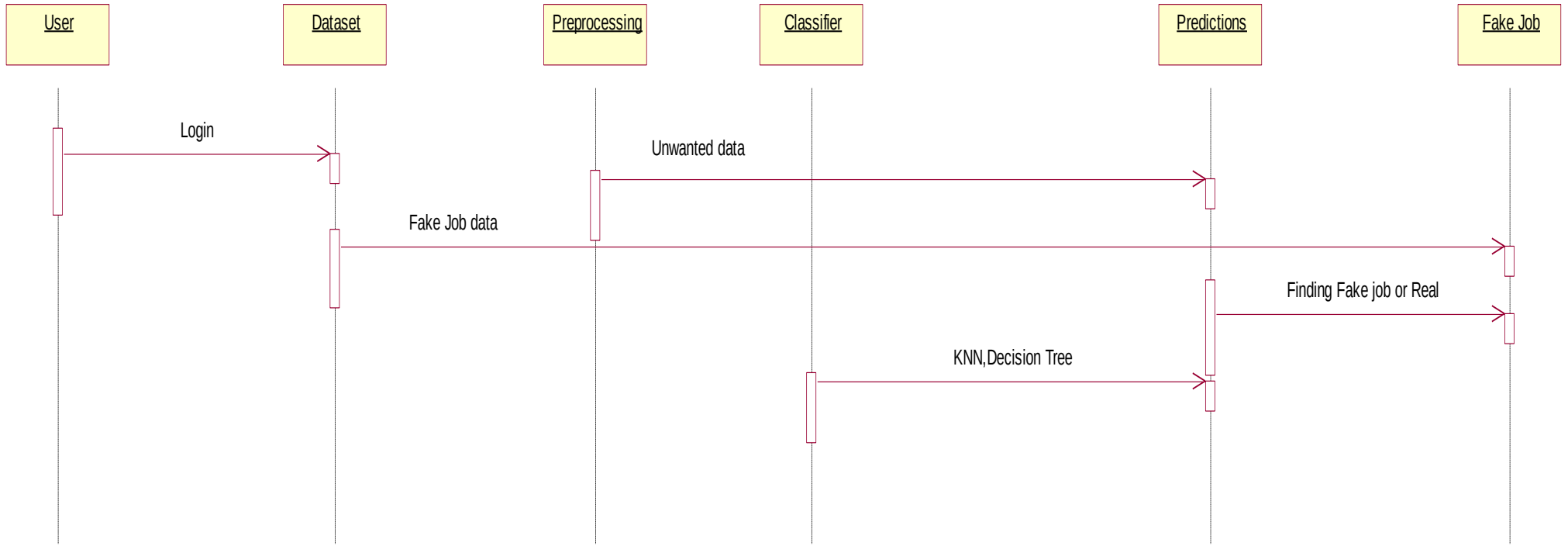
- Once you’re confident enough to take your trained and tested model into the production-ready environment,

- the first step is to save it into a .h5 or. pkl file using a library like pickle

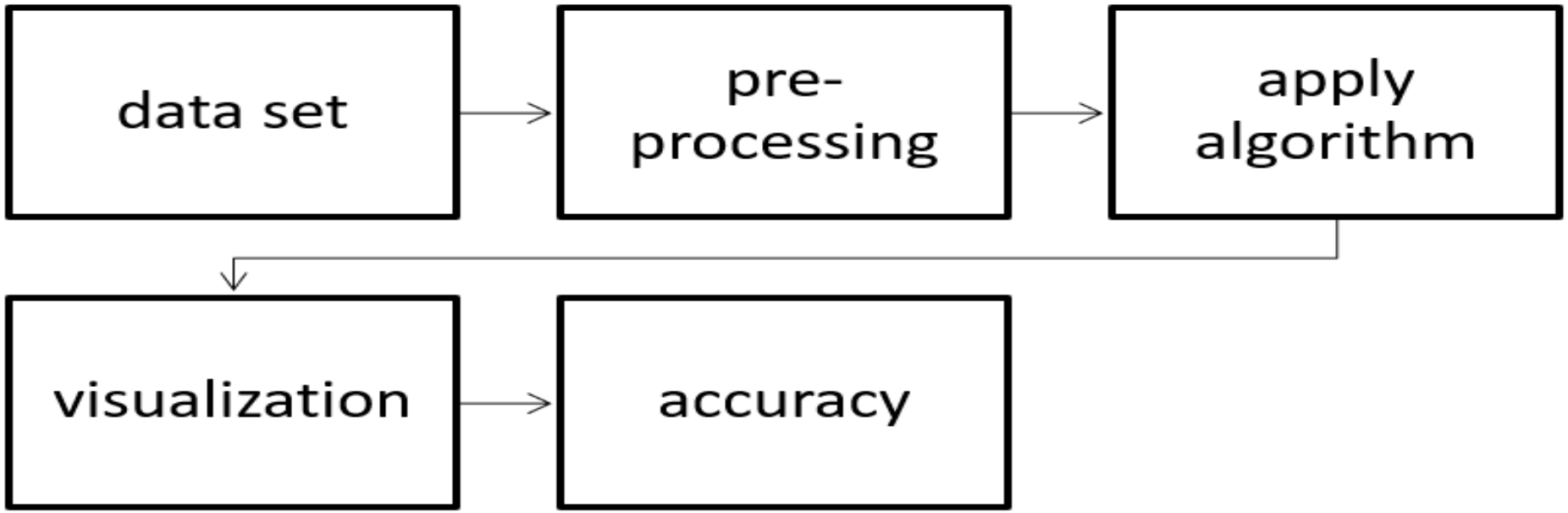
USE CASE DIAGRAM



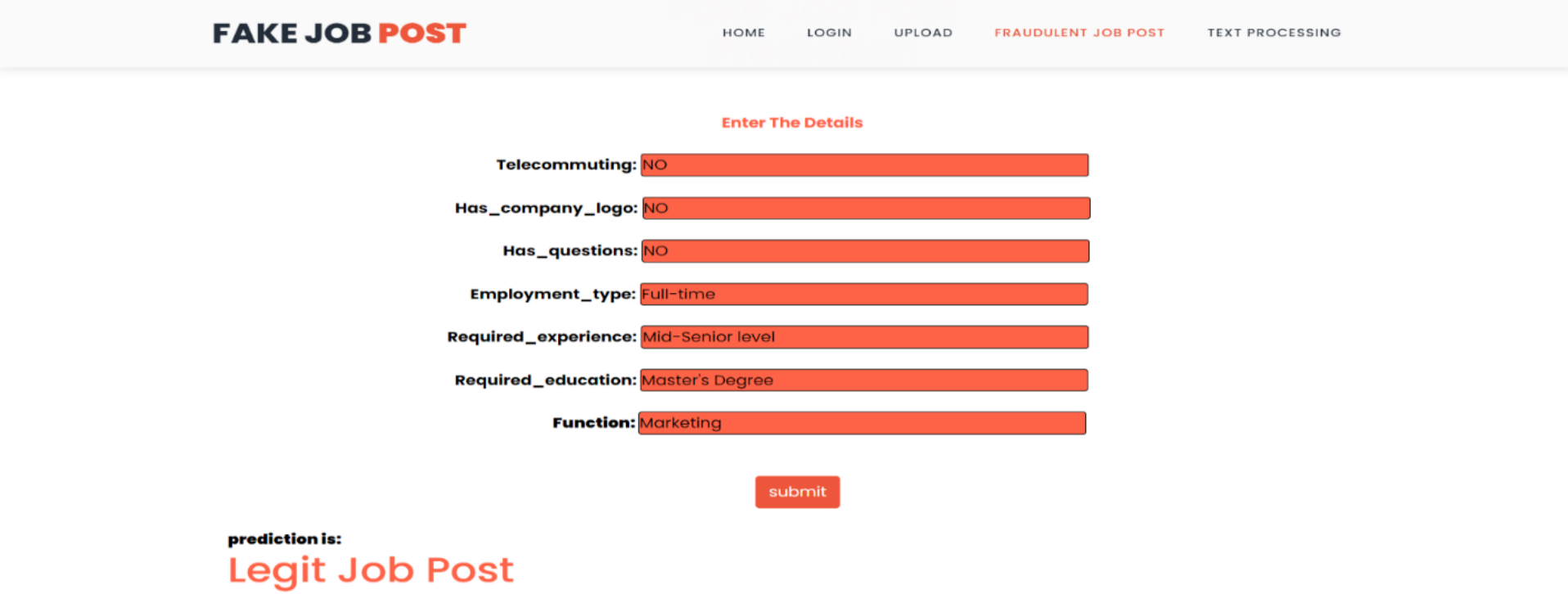
SEQUENCE DIAGRAM



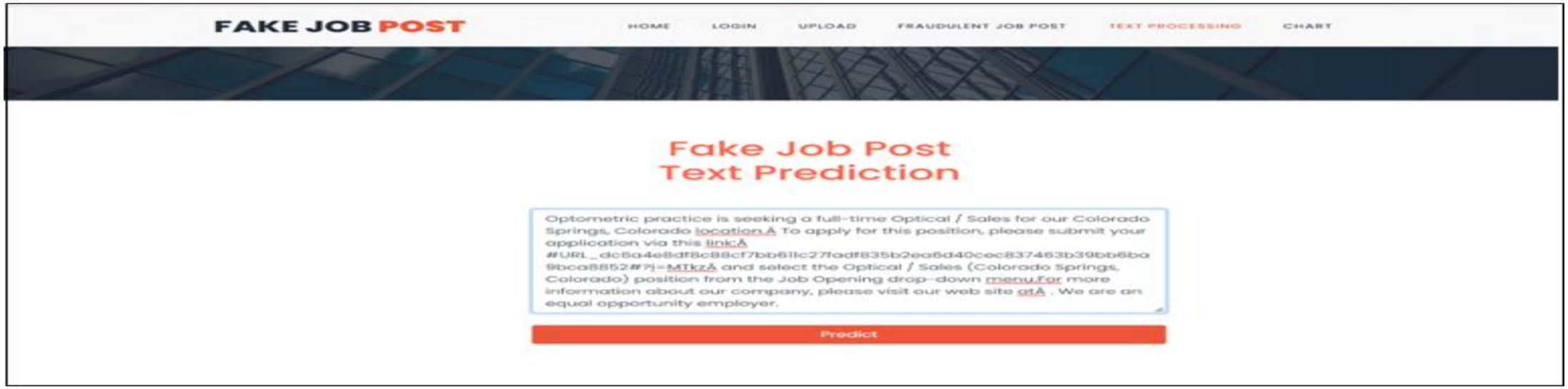
DATA FLOW DIAGRAM



SNAPSHOTS


FAKE JOB POST
HOME
LOGIN
UPLOAD
FRAUDULENT JOB POST
TEXT PROCESSING
Enter The Details
Telecommuting: NO
Has_company_logo: NO
Has_questions: NO
Employment_type: Full-time
Required_experience: Mid-Senior level
Required_education: Master's Degree
Function: Marketing
submit
prediction is:
Legit Job Post

SNAPSHOTS


FAKE JOB POST
HOME
LOGIN
UPLOAD
FRAUDULENT JOB POST
TEXT PROCESSING
CHART
Fake Job Post
Text Prediction
Optometric practice is seeking a full-time Optical / Sales for our Colorado Springs, Colorado location.Â To apply for this position, please submit your application via this link:Â
#URL_dc5a4e8df8c88cf7bb61lc27fadf835b2ea6d40cec837463b39bb6ba9bca8852#?i=MTkzÂ and select the Optical / Sales (Colorado Springs, Colorado) position from the Job Opening drop-down menu.For more information about our company, please visit our web site atÂ . We are an equal opportunity employer.
Predict


FAKE JOB POST
HOME
LOGIN
UPLOAD
FRAUDULENT JOB POST
TEXT PROCESSING
CHART
Fake Job Post
Text Prediction
Enter The Details!
Predict
prediction is:
Fake Job Post

FUTURE ENHANCEMENT

- we intend to expand EMSCAD and enrich the ruleset by focusing on user behavior, company and network data as well as user-content-IP collision patterns.
- Moreover, we would like to employ graph modeling and explore connections between fraudulent job ads, companies, and users.
- Ultimately, our goal is to propose an applicable employment fraud detection tool for commercial purposes

CONCLUSION

- Fake job detection has become a great concern all over the world at present we have analyzed the impacts of fake job posts, to detect fake job posts we have used Random Forest, DNN classifiers which gives the 98% classification accuracy.
- We have implemented with EMSCAD dataset which contains real life fake job posts.