

PROFESSIONAL TRAINING REPORT
at
Sathyabama Institute of Science and Technology
(Deemed to be University)

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Technology in Information Technology

By

LAKSHMI PRIYA M
REG. NO. 38120044



DEPARTMENT OF INFORMATION TECHNOLOGY

SCHOOL OF COMPUTING

SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600119, TAMILNADU

NOVEMBER 2021



SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade “A” by NAAC
(Established under Section 3 of UGC Act, 1956)
JEPPIAAR NAGAR, RAJIV GANDHI SALAI
CHENNAI– 600119
www.sathyabama.ac.in



DEPARTMENT OF INFORMATION TECHNOLOGY

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **LAKSHMI PRIYA M**
(Reg.No: 38120044) who carried out the project entitled “**Real and Fake Job Prediction**” under my supervision from June 2021 to November 2021.

Internal Guide

Dr.V.Nirmalrani M.E.Ph.D

Head of the Department

Dr.R.Subhashini M.E.,Ph.D.,

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I, **LAKSHMI PRIYA M** hereby declare that the project report entitled **Real and Fake Prediction** done by me under the guidance of **Dr.v.Nirmalrani M.E.,Ph.D.**, is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr.T.Sasikala M.E., Ph.D, Dean, School of Computing, Dr.R.Subhashini, M.E., Ph.D. Head of the Department of Information Technology** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr.v.Nirmalrani M.E., Ph.D.,** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Information Technology** who were helpful in many ways for the completion of the project.

Real and fake Job posting using machine learning technique

Abstract:

Everything is being online now it allows people to reduce their manual efforts all of the job postings are posted online now so it gives the company a wide range of area to gather the talented candidates and for the people who are searching they can also know the information most companies can directly post the job. All job postings which are posted are not true there are fraudulent job postings also. So we try to classify the fraudulent posting from real. The aim is to predict machine learning based techniques for real or fake job prediction results in best accuracy. The analysis of dataset is done by supervised machine learning technique(SMLT) to capture several information's like, variable identification, uni-variate analysis, bi-variate and multi-variate analysis, missing value treatments and analyze the data validation, data cleaning/preparing and data visualization will be done on the entire given dataset

Keywords:dataset, Machine learning-Classification method, python, Prediction of Accuracy result.

Objectives:

The goal is to develop a machine learning model for real-time air quality forecasting, to potentially replace the updatable supervised machine learning classification models by predicting results in the form of best accuracy by comparing supervised algorithm.

Existing System

A mathematical model to study the dynamic spreading and controlling activities of message transmission in OSN was proposed. The proposed model employs differential equations for investigating the effect of verification and blocking of users and the spread of messages on OSNs. It describes how misinformation gets disseminated among groups with the influence of different misinformation refuting measures. This model is based on different types of epidemic classes and has two layers of control mechanism to control the rumor in the social network. This model assumes that all users are susceptible that means anyone may turn a victim of misinformation or untrusted message. For protection, initially, the users are authenticated using a verified class. Hence, before accepting the request of any user, the user authentication method is applied, and the reliability of the messages from this user is evaluated in order to minimize the activities of malicious users to the OSN. If the value of R_0 is less than one ($R_0 < 1$), then fake message spreading in the online network will not be prominent, otherwise if $R_0 > 1$ the rumor will persist in the OSN.

Drawbacks:

- This is just a mathematical model it does not classify the fake news.
- Accuracy, Recall, F1 score metrics are not calculated and machine learning algorithms are not applied.

Proposed System:

The proposed method is built a machine learning model to classify the real or fake job posting to overcome this method to implement machine learning approach by user interface of GUI application. The dataset is first preprocessed and the columns are analyzed to see the dependent and independent variable and then different machine learning algorithms would be applied to extract patterns and to obtain results with maximum accuracy.

Exploratory Data Analysis

Multiple datasets from different sources would be combined to form a generalized dataset, and then different machine learning algorithms would be applied to extract patterns and to obtain results with maximum accuracy.

Data Wrangling

In this section of the report will load in the data, check for cleanliness, and then trim and clean given dataset for analysis. Make sure that the document steps carefully and justify for cleaning decisions.

Data collection

The data set collected for predicting given data is split into Training set and Test set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model which was created using machine learning algorithms are applied on the Training set and based on the test result accuracy, Test set prediction is done.

Building the classification model

It provides better results in classification problem.

- It is strong in preprocessing outliers, irrelevant variables, and a mix of continuous, categorical and discrete variables.
- It produces out of bag estimate error which has proven to be unbiased in many tests and it is relatively easy to tune with.

Environmental Requirements:

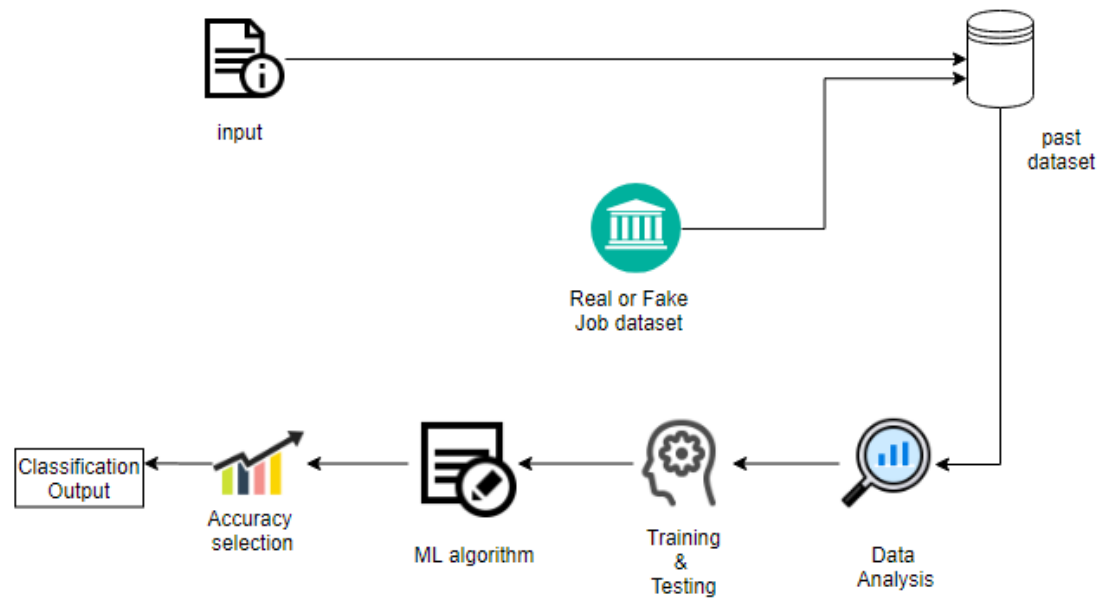
1. Software Requirements:

Operating System	: Windows
Tool	: Anaconda with Jupyter Notebook

2. Hardware requirements:

Processor	: Pentium IV/III
Hard disk	: minimum 80 GB
RAM	: minimum 2 GB

System Architecture



Future Work

- To automate this process by show the prediction result in web application or desktop application.
- To optimize the work to implement in Artificial Intelligence environment.

Code

```
"cells":[
  {
    "cell_type":"markdown",
    "metadata":{},
    "source":[
      "#BuildamodelforLSTMNeuralNetwork"
    ]
  },
  {
    "cell_type":"code",
    "execution_count":1,
    "metadata":{},
    "outputs":[],
    "source":[
      "importwarnings\n",
      "warnings.filterwarnings(\"ignore\")"
    ]
  },
  {
    "cell_type":"code",
    "execution_count":2,
    "metadata":{"colab":{"base_uri":"https://localhost:8080/", "height":284}},
    "colab_type":"code",
    "id":"SuVP4Wf8SZUG",
    "outputId":"f09a2e7b-225a-4883-d3ca-4cdf901460dd"
  },
  {
    "outputs":[],
    "source":[
      "#!pipinstallkeras-metrics"
    ]
  },
  {
    "cell_type":"code",
    "execution_count":3,
    "metadata":{"colab":{}},
    "colab_type":"code",
    "id":"gDTF5giuRZ73"
  },
  {
    "outputs":[
      {
        "name":"stderr",
        "output_type":"stream",
        "text":[
          "UsingTensorFlowbackend.\n"
        ]
      }
    ],
    "source":[
      "importtqdm\n",
      "importnumpyasnp\n",
      "importkeras_metrics#forrecallandprecisionmetrics\n",
```

```

"fromtensorflow.keras.preprocessing.sequenceimportpad_sequences\n",
"fromtensorflow.keras.preprocessing.textimportTokenizer\n",
"fromtensorflow.keras.layersimportEmbedding,LSTM,Dropout,Dense\n",
"fromtensorflow.keras.modelsimportSequential\n",
"fromtensorflow.keras.utilsimportto_categorical\n",
"fromtensorflow.keras.callbacksimportModelCheckpoint,TensorBoard\n",
"fromsklearn.model_selectionimporttrain_test_split\n",
"importtime\n",
"importnumpyasnp\n",
"importpickle\n",
"\n",
"fromkeras.preprocessing.textimportTokenizer"
]
},
{
"cell_type":"code",
"execution_count":4,
"metadata":{
"colab":{},
"colab_type":"code",
"id":"xKcUd3fZShHe"
},
"outputs":[],
"source":[
"SEQUENCE_LENGTH=100#thelengthofallsequences(numberofwordspersample)\n",
"EMBEDDING_SIZE=100#Using100-DimensionalGloVeembeddingvectors\n",
"TEST_SIZE=0.20#ratiooftestingset\n",
"\n",
"BATCH_SIZE=64\n",
"EPOCHS=30#numberofepochs\n",
"\n",
"#toconvertlabelstointegersandvice-versa\n",
"#label2int={\"truthful\":0,\"deceptive\":1}#added\n",
"#int2label={0:\"truthful\",1:\"deceptive\"}#added\n",
"\n",
"label2int={\"real\":0,\"fraudulent\":1}#added\n",
"int2label={0:\"real\",1:\"fraudulent\"}#added"
]
},
{
"cell_type":"code",
"execution_count":5,
"metadata":{
"colab":{}
},
"outputs":[],
"source":[
"importpandasaspd"
]
},
{
"cell_type":"code",
"execution_count":6,
"metadata":{
"colab":{}
},
"outputs":[],
"source":[
"data=pd.read_csv(\"fake_job_postings.csv\")"
]
},
{
"cell_type":"code",

```

```

"execution_count":7,
"metadata":{},
"outputs":[
{
"data":{
"text/plain":[
"job_id0\n",
"title0\n",
"location346\n",
"department11547\n",
"salary_range15012\n",
"company_profile3308\n",
"description1\n",
"requirements2695\n",
"benefits7210\n",
"telecommuting0\n",
"has_company_logo0\n",
"has_questions0\n",
"employment_type3471\n",
"required_experience7050\n",
"required_education8105\n",
"industry4903\n",
"function6455\n",
"fraudulent0\n",
"dtype:int64"
]
},
"execution_count":7,
"metadata":{},
"output_type":"execute_result"
},
"source":[
"data.isnull().sum()"
],
{
"cell_type":"code",
"execution_count":8,
"metadata":{},
"outputs":[
"source":[
"df=data.dropna()"
]
},
{
"cell_type":"code",
"execution_count":9,
"metadata":{},
"outputs":[
{
"data":{
"text/plain":[
"job_id0\n",
"title0\n",
"location0\n",
"department0\n",
"salary_range0\n",
"company_profile0\n",

```

```

"description0\n",
"requirements0\n",
"benefits0\n",
"telecommuting0\n",
"has_company_logo0\n",
"has_questions0\n",
"employment_type0\n",
"required_experience0\n",
"required_education0\n",
"industry0\n",
"function0\n",
"fraudulent0\n",
"dtype:int64"
],
},
"execution_count":9,
"metadata":{},
"output_type":"execute_result"
},
],
"source":[
"df.isnull().sum()"
],
},
{
"cell_type":"code",
"execution_count":10,
"metadata":{},
"outputs":[
],
"source":[
"x=df['description']"
],
},
{
"cell_type":"code",
"execution_count":11,
"metadata":{},
"outputs":[
{
"data":{
"text/plain":[
"6YourResponsibilities:聽ManagetheEnglish-spea...\n",
"15AboutVaultDragon聽VaultDragonisDropboxfor...\n",
"23#URL_eda2500ddcedb60957fcd7f5b164e092966f8c4e8...\n",
"98IC&ETechnician|Bakersfield,CAMt.Poso...\n",
"102Thejobistosupportthegrowthofthe#URL_9...\n",
"... \n",
"17734WeareSeekingacandidatewhosecorevaluesi...\n",
"17759WeareSeekingacandidatewhosecorevaluesi...\n",
"17813WeareSeekingacandidatewhosecorevaluesi...\n",
"17849BalancedLabsexiststoimprovethe livesofa...\n",
"17865AvenueStoryis聽lookingforacurious,ambitiou...\n",
"Name:description,Length:774,dtype:object"
]
}
},
"execution_count":11,
"metadata":{},
"output_type":"execute_result"

```

```

}
],
"source":[
"x"
]
},
{
"cell_type":"code",
"execution_count":12,
"metadata":{},
"outputs":[],
"source":[
"x=list(x)"
]
},
{
"cell_type":"code",
"execution_count":13,
"metadata":{},
"outputs":[
{
"data":{
"text/plain":

```

REFERENCES

- [1] F. Ferdiansyah and J. S. Suroso, "Evaluation of knowledge management system to improve the performance of employees at PT Data Citra Mandiri," 2017 International Conference on Applied Computer and Communication Technologies (ComCom), 2017, pp. 1-6, doi:10.1109/COMCOM.2017.8167086.
- [2] H. Liu and X. Liu, "A Study on Knowledge Employees Based on Management optimization Strategies," 2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), 2019, pp. 443-447, doi: 10.1109/ISKE47853.2019.9170308.
- [3] H. Juwitasary, L. Christian, E. P. Putra, F. Filia and W. Sardjono, "Employee Process Management System Implementation Model for Improving Employee Performance," 2018 International Conference on Information Management and Technology

(ICIMTech), 2018, pp. 367-372, doi:10.1109/ICIMTech.2018.8528135.

- [4] M. Paandian and M. B. Malarvili, "A Web based Employee Medical History Management and Monitoring System," 2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences, 2012, pp. 789-794, doi:10.1109/IECBES.2012.6498113.