APPENDIX 1

PERSONALITY PREDICTION THROUGH

CV ANALYSIS

END TERM REPORT

by

NAME OF THE CANDIDATE(S): T. SINDHU.(01)
B.LOHITHA.(17)
SUDARSHAN.(18)
ANIL KUMAR.(47)

SECTION: K18HV

Department of Intelligent Systems

School of Computer Science Engineering

Lovely Professional University, Jalandhar.

APRIL-2020.

APPENDIX 2

Student Declaration

This is to declare that this report has been written by me/us. No part of the report is copied from other sources. All information included from other sources have been duly acknowledged. I/We aver that if any part of the report is found to be copied, I/we are shall take full responsibility for it.

Signature of Student :
Name of Student: T. Sindhu
Roll number: 01

Signature of Student
Name of Student: B.Lohitha
Roll number:17

Signature of Student
Name of Student: Sudarshan
Roll number:18

Signature of Student
Name of Student:Anil kumar
Roll number:47

Place :

Date :

APPENDIX 3

TABLE OF CONTENTS

TITLE   **PAGE NO.**

APPENDIX 4

BONAFIDE CERTIFICATE

Certified that this project report "**PERSONALITY PREDICTION THROUGH CV ANALYSIS**" is the bonafide work of "**T. sindhu , B. Lohitha, Sudarshan, Anil Kumar**" who carried out the project work under my supervision.

<<Signature of the Supervisor>>

<<Name of supervisor>>

<<Academic Designation>>

<<ID of Supervisor>>

<<Department of Supervisor>>

1.BACKGROUND AND OBJECTIVE OF PROJECT ASSIGNED

Today in many of the companies of the world the most difficult task in recruitment of new employee is to check everyones resume and take out the candidates who are eligible for certain post.It requires a lot of time and workforce.Here a comes the need of a system which can automatically match the candidates resume with the job and then it can be very easy for the companies to get there required candidates for the certain job.

So to make this system we need to know to some key things like natural language processing and spacy module in python etc.We also need to take a bunch of cvs to import it in the code by giving the path of the cv folder so the system can parse the cvs and take out the fittest cvs using matching with certain keywords in cv.

This project parses the whole resume and then it matches using the keywords and it makes a csv file who all are fit for the job .After that using that data the system plots a bar graph using which we can accurately predict who is best for the job.This project is very essential in this fast growing world as it reduces the burden which caused in companies for recruitment of new employes .

For an example lets assume a company wants to get all the employes who are from data science field.Now the company will put all the resmues in a drive and put the path in the code and the system will parse the whole resume and take out all the words related to datascience like NLP or machine learning and then it will create a csv file with all the names who are fit for it and then for accuracy it will create a bra graph to get the best employes.

2.DESCRIPTION   OF PROJECT:-

First   of all we have to give the   path of   all the resumes in the code.
mypath=`'D:/NLP_Resume/Candidate'`

Next to read the resumes we have to use pdf extractor.We will use
pypdf2.pdfreader function.

These are the keywords used to take out the data science resumes:-

Statistics	Machine Learning	Deep Learning	R Language	Python Language	NLP	Data Engineering
statistical models	linear regression	neural network	r	python	nlp	aws
statistical modeling	logistic regression	keras	ggplot	flask	natural language processing	ec2
probability	K means	theano	shiny	django	topic modeling	amazon redshift
normal distribution	random forest	face detection	cran	pandas	lda	s3
poisson distribution	xgboost	neural networks	dplyr	numpy	named entity recognition	docker
survival models	svm	convolutional neural network (cnn)	tidyr	scikitlearn	pos tagging	kubernetes
hypothesis testing	naïve bayes	recurrent neural network(RNN)	lubridate	sklearn	word2vec	scala
bayesian inference	pca	object detection	knitr	matplotlib	word embedding	teradata
factor analysis	decision trees	yolo		scipy	lsi	google big query
forecasting	svd	gpu		bokeh	spacy	aws lambda
markov chain	ensemble models	cuda		statsmodel	gensim	aws emr
monte carlo	boltzman machine	tensorflow			nltk	hive
		lstm			nmf	hadoop
		gan			doc2vec	sql
		opencv			cbow	
					bag of words	
					skip gram	
					bert	
					sentiment analysis	
					chat bot	

I have mainly used these keywords to take out data science
candidates from all the resumes .

Now after matching the system creates a csv file of all the
candidates that are of data science background.

Candidate Names	Keyword Category						
	DE	DL	ML	NLP	Python	R	Stats
ariadne page	2	1	2	2	2	1	2
arthur levitt	3	2	3	2	1	1	1
dom cobb	2	5	4	10	5	1	0
fischer	0	0	5	0	0	0	8
leo cap	1	5	6	1	2	1	0
sato	5	0	4	1	1	2	2
yusuf	5	2	4	5	1	2	2

Now atlast to create more clarity system will plot a bar graph using this data by which we can predict all candidates who are eligible.

Resume keywords by category
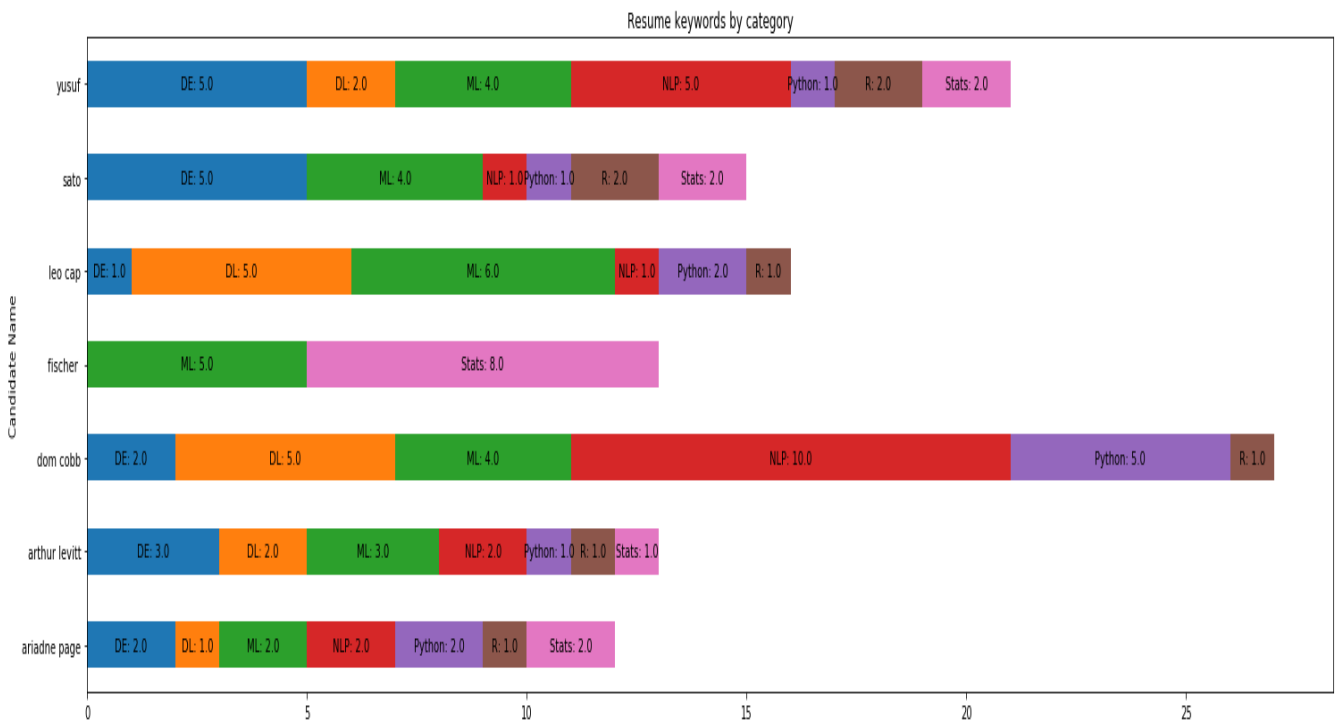
Now   from all the resumes   uploaded we are having all the candidates who are from data science resume.

DESCRIPTION   OF   ROLES   AMONG   STUDENTS :-

1.Sudarshan kumar Thakur(11802596):-Studied about all the libraries required to make this project and   coded the resume reading and parsing   part in the code.

2.LOHITHA(11801494):-Studied about pandas and matplotlib   libraries and coded the plotting the graph.

3.Sindhu(11804718):-Studied about all the python libraries arranged all the cvs   data and coded the keyword matcher part and made the report .

4.Anil(11803696):-Studied about nlp and spacy and worked on it and helped   in   preparaing report.

Technologies and Frameworks used in projects:-

1.Python idle   or any python compiler
2.Spacy library   for natural Languge processing
3.Pdfextractor to read resumes
4.Matplotlib for plotting the graph

8. SWOT Analysis achieved in project.

Strengths:

1. To provide a more effective way of short-listing the candidates through their resumes.

2. System will rank the experience and key skills required for particular job position.

3. Bar-graph is plotted in which, we can select the desired candidate more accurately.

4. To produce ranking decisions that would have relatively higher consistency than those of human experts.

Weakness :

1. This system requires large memory space as it stores data related to CV's.

2. Requires an active internet connection.

3. May provide inaccurate results if data not entered properly.

Opportunities:

1. This system can be used in many business sectors that may require expert candidate.

Threats:

1. Resumes may be leaked.

2. Security threats.

3. We can lose our resumes if the system crash happens.