HUMAN RESOURCES ANALYSIS SYSTEM

A MINI PROJECT REVIEW

Submitted by

ADITYA S NAIR   (720419104001)

AMAL R   (720419104002)

KARTHIK M   (720419104012)

SANJAY RJ   (720419104022)

VINAYAK PK   (720419104031)

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

CMS COLLEGE OF ENGINEERING AND TECHNOLOGY COIMBATORE

ANNA UNIVERSITY: CHENNAI 600 025

JUNE 2022

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **"HUMAN RESOURCES ANALYSIS SYSTEM"** is the bonafide work of "**ADITYA S NAIR, VINAYAK PK, AMAL R, SANJAY RJ, KARTHIK**" who carried out the project work under my supervision.

SIGNATURE   **SIGNATURE**

Dr.G.Chitra Ganapathy,   **Dr.G.Chitra Ganapathy,**

HEAD OF THE DEPARTMENT   **GUIDE**

Professor ,   **Professor ,**

Department of Computer Science & Engineering,

Department of Computer Science & Engineering,

CMS College of Engineering and Technology,

CMS College of Engineering and Technology,

Coimbatore-641 032   Coimbatore-641 032

Submitted for the Anna University Project Viva voice held on

.

INTERNAL EXAMINER   **EXTERNAL EXAMINAR**

ACKNOWLEDGEMENT

Any organized and systematic word calls for the corporation of team of people. Our project does not have any exception to this. Hence these pages find the space for thanking all those who have directly and indirectly contributed to completion of this work in a successful manner.

We record our indebtedness to our **Principal Dr.N.Sudha,M.E,Ph.D** for her guidance and sustained encouragement for the successful completion of this project.

We express our heartiest thanks to **Dr.G.Chitra Ganapathy,M.E,Ph.D, Professor, Head Of The Department**, Department of Computer Science and Engineering ,CMS College of Engineering and Technology, for his encouragement and valuable guidance in carrying out our project work.

We express our heartfelt thanks to Project Guide **Dr.G.Chitra Ganapathy,M.E,Ph.D, Professor**, CMS College of Engineering and Technology, Coimbatore, for her valuable and timely support for our project work.

We express our heartfelt thanks to Project Co-ordinator **Mr.S.Dinesh Kumar, M.E.**, **Assistant Professor**, CMS College of Engineering and Technology, Coimbatore, for her valuable and timely support for our project work.

We also express thanks to our parents, friends for their encouragement and best wishes in the successful completion of this dissertation.

ABSTRACT

The economic environment has changed dramatically around the world in recent years, generating favorable conditions for the growth of small- and medium-sized firms. The socioeconomic development and international integration of China are greatly influenced by the growth in both quality and quantity, the scale of operations, and the internal force of small- and medium-sized businesses. Moreover, in comparison with other developed countries around the world, Chinese small- and medium-sized enterprises continue to face many limitations in terms of size and contribution levels and have not yet fully realized their potentials due to difficulties and poor quality; human resources in this field are still lacking. This study defines the current state of human resources in small and medium firms, the factors that impede development, and the steps that can be taken to overcome these obstacles in order to assist human resource development in this sector during the current period. This study uses machine learning (ML) techniques to manage and analyze human resource data in modern enterprises. The ML techniques realize the functions of the human resource system and reduce the business volume in human resource in order to improve the efficiency and management of the human resource work. In this paper, we designed and implemented the wage forecasting model in human resources that uses a gradient descent algorithm, its types, and backpropagation (BP) neural network to improve the accuracy of the forecasting model. We performed multiple experiments by using a various number of neurons in the hidden layers, different number of iterations, and several types of gradient descent algorithms. The BP neural network model was performed brilliantly by attaining the training accuracy of 89.98% and validation accuracy of 84.05%. The experimental results show the significance and importance of the proposed work.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The problem is that the present are not much flexible and efficient and time saving. It requires candidate, to fill the forms online than also you might not get the genuine information of the candidate. Beside

Where our system which saves the time of the candidate by providing to upload there resume in any format preferable to the candidate beside all the information in the resume our system will detect all its activity from the candidate social profile which will give the best candidate for that particular job and candidate will also be satisfied because he will get job in that company which really appreciates candidates skill and ability. On the other hand we are providing same kind of flexibility to the client company.The current recruitment process are more tedious and time consuming which forces the candidates to fill all their skill and information manually. And HR team requires more man power to scrutinize the resumes of the candidates. So that motivated to build a solution that is more flexible and automated.

The major objective of our system is to take the current resume ranking system to other level and makes it more flexible for both the entity.

1)   Candidates, who has been hired.

2)   Client company, who is hiring the candidates.

Candidates, who has been hired :

Candidates who are searching for jobs after been graduated. Out of those, major number of candidates are so much desperate that they are ready to work on any post irrelevant to their skill set and ability. The main reason behind this unemployment is like a cancer to our society, if a guy is not got place after been passed out for 1yr, society include relatives starting blaming that guy. Inspite of this reason the candidate are ready to work in any condition, on any post. So they don't have to face those situation.

Where our system help such candidates to get hired by such a company or an organisation who really worth their ability and their skill sets. Where our algorithm will work in such a way that with the help of the previous result and previous ranking constraints, it will try to optimize the current result, which we called it Machine Learning.

This will make sure that the relevant candidate is been hired for that particular vacancy.

CHAPTER 2

LITERATURE SURVEY

2.1 Research and application of practice teaching reform of HRM major under the background of AI

By: Tian Wang, Jianbang Lin

Artificial intelligence (AI) is based on manpower. It is not only intelligence and machines, which highlight the contribution of advanced nature and innovation to the intelligence, so as to serve human resource management better. Human resource management (HRM) is people-oriented, emphasizing on the actual situation of people and adopting different ways to achieve the management effect. Recognizing the relationship between AI and HRM is the key problem that HRM practitioners need to solve in planning their career. The development of AI challenges the work of HRM, and at the same time puts forward higher requirements for the training of HRM professionals. Facing the impact of science and technology, the practical teaching reform of HRM major is urgent. Based on the belief that reform is necessary, colleges and professional teachers should break through the inherent cognition, work hard on the reform measures and contents, and create the HRM major from the perspective of innovation.

2.2 Experimental Verification on Human-Centric Network-Based Resource Allocation Approaches for Process-Aware Information Systems

By: Moon-Sook Yeon, Young-Koo Lee, Dinh-Lam Pham, Kwanghoon Pio Kim

Efficient allocating human resource obtains significant advantages to business operations in enterprises. Manipulating the available human resources of a business optimizes the expense of developing products and services, improves work productivity, and brings more value to the business. In the process of creating products and benefits in the enterprise, a network of working relationships is formed among employees in the organization. In this paper, we propose an approach to support decision-making on human resource allocation in an organization based on the human-centric network discovered from the event logs. To be more precise, the knowledge derived from human-centric networks formed through the operation phases of the information systems is used to make resource allocation strategies in the organizations. Using the appropriateness probabilities and social network metrics for allocating human resources, we clarify efficiently allocating resources based on different metrics. The system architecture and the scenarios are also described for planning human resources allocation based on these networks. Furthermore, we use real-life data set for getting experimental verification and proving the proposal.

2.3 A Literature Review on application of Artificial Intelligence in Human Resource Management and its practices in current organizational scenario

By: Gaurav Sharma

Artificial Intelligence simplifies people's job by eliminating repeated tasks and providing unbiased and valuable insights. People have a perception that artificial intelligence will replace human efforts and can be the reason for mass termination of human resources. According to one research, 71% of companies see HR analytics as a high priority in their organizations. Also 8% of organizations report that they have usable data.The present paper studies the concept of Artificial Intelligence (AI) and its application on various human resource dimensions. For that, a conceptual framework is also given depicting how Artificial Intelligence (AI) benefits HR and its various dimensions. Based on an extensive literature review, this paper will discuss the use of best practices of Artificial Intelligence (AI) in Human resource functions. HR analytics is also considered a main component of Artificial Intelligence (AI) in HR practices.

CHAPTER 3
MODULES

3.1 MODULE DESCRIPTION

MODULES

 • DATA COLLECTION

 • DATA PRE-PROCESSING

• FEATURE EXTRATION

• EVALUATION MODEL

Module Description:

DATA COLLECTION

Data used in this paper is a set of student details in the school records. This step is concerned with selecting the subset of all available data that you will be working with. ML problems start with data preferably, lots of data (examples or observations) for which you already know the target answer. Data for which you already know the target answer is called labelled data.

DATA PRE-PROCESSING

Organize your selected data by formatting, cleaning and sampling from it.

Three common data pre-processing steps are:

1. Formatting

 2. Cleaning

 3. Sampling

Formatting: The data you have selected may not be in a format that is suitable for you to work with. The data may be in a relational database and you would like it in a flat file, or the data may be in a proprietary file format and you would like it in a relational database or a text file.

Cleaning: Cleaning data is the removal or fixing of missing data. There may be data instances that are incomplete and do not carry the data you believe you need to address the problem. These instances may need to be removed. Additionally, there may be sensitive information in some of the attributes and these attributes may need to be anonym zed or removed from the data entirely.

Sampling: There may be far more selected data available than you need to work with. More data can result in much longer running times for algorithms and larger computational and memory requirements. You can take a smaller representative sample of the selected data that may be much faster for exploring and prototyping solutions before considering the whole dataset.

3.2 FEATURE EXTRATION

Next thing is to do Feature extraction is an attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes. The transformed attributes, or features, are linear combinations of the original attributes. Finally, our models are trained using Classifier algorithm. We use classify module on Natural Language Toolkit library on Python.We use the labelled dataset gathered. The rest of our labelled data will be used to evaluate the models. Some machine learning algorithms were used to classify pre-processed data. The chosen classifiers were Random forest. These algorithms are very popular in text classification tasks.

3.3 EVALUATION MODEL

Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it can easily generate overoptimistic and over fitted models. There are two methods of evaluating models in data science, Hold-Out and Cross Validation to avoid over fitting, both methods use a test set (not seen by the model) to evaluate model performance. Performance of each classification model is estimated base on its averaged. The result will be in the visualized form. Representation of classified data in the form of graphs. Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

While working with data, it can be difficult to truly understand your data when it's just in tabular form. To understand what exactly our data conveys, and to better clean it and select suitable models for it, we need to visualize it or represent it in pictorial form. This helps expose patterns, correlations, and trends that cannot be obtained when data is in a table or CSV file.

The process of finding trends and correlations in our data by representing it pictorially is called Data Visualization. To perform data visualization in python, we can use various python data visualization modules such as Matplotlib, Seaborn, Plotly, etc. In this article, The Complete Guide to Data Visualization in Python, we will discuss how to work with some of these modules for data visualization in python and cover the following topics in detail.

- Data Visualization in Python

- Matplotlib and Seaborn

- Line Charts

- Bar Graphs

- Histograms

- Scatter Plots

- Heat Maps

CHAPTER 4

SYSTEM ANALYSIS

4.1 EXISTING SYSTEM

The existing systems or Prior systems needed lot of human efforts

 time.Cost of hiring is high.

4.1.1 DRAWBACK OF EXISTING SYSTEM

1.Potential candidate may loose the opportunity because of ambigious keyword matching.

2.Resumes needed to be in specific format

4.2 PROPOSED SYSTEM

1.   Use of NLP to read resumes allow candidates the freedom to choose any format that's avalible to them.

2.   Machine learning is used to rank candidates in accordiance to requirements Which reduces the efforts of sorting thousands of resumes.

4.2.1 ADVANTAGE OF PROPOSED SYSTEM

1.   Use of NLP can be used to get mean out of ambigius data.

2.   Five benifits of A.I.

Goes Beyond Key Words

- Fast and Accurate
- Perfect For the New World of Social Recruiting
- Customizes to your Need
- Get Smarter

CHAPTER 5
SYSTEM REQUIREMENTS

5.1 HARDWARE SPECIFICATION

- ❖ Processor: Intel Pentium 4
- ❖ Speed: 1.2 GHz
- ❖ RAM: 256MB
- ❖ Hard disk: 80GB
- ❖ Monitor: 17 inch

5.2 SOFTWARE SPECIFICATION

- ❖ Front end: Jupyter using PYTHON
- ❖ Back end: DataSet(CSV)
- ❖ Operating system: Windows XP

5.2.1 ABOUT THE SOFTWARE

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Often, programmers fall in love with Python because of the increased productivity it provides. Since there is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging Python programs is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception. When the program doesn't catch the exception, the

interpreter prints a stack trace. A source level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping through the code a line at a time, and so on. The debugger is written in Python itself, testifying to Python's introspective power. On the other hand, often the quickest way to debug a program is to add a few print statements to the source: the fast edit-test-debug cycle makes this simple approach very effective.

• Python 3.7: Python is an interpreted, high level, general programming

• language. Its formatting is visually uncluttered, and it often uses English

keywords where other languages use punctuation. It provides a vast library for data mining and predictions.

• Jupiter Notebook/ Spider/ PyCharm: It is an open source cross-platform integrated development environment (IDE) for scientific programming in the Python language. Spyder integrates with a number of prominent packages as well as another open-source software.

• NumPy: NumPy was used for building the front-end part of the system.

• Pandas: Pandas was used for the data pre-processing and statistical analysis of data.

• Matplotlib: Matplotlib was used for the graphical representation of our prediction.

CHAPTER 6

SYSTEM TESTING AND IMPLEMENTATION

6.1 SYSTEM TESTING

Testing methodologies:

THE TERM SYSTEM TESTING CAN BE USED IN A NUMBER OF WAYS. IN A GENERAL SENSE, THE TERM 'SYSTEM TESTING' REFERS TO THE TESTING OF THE SYSTEM IN ARTIFICIAL CONDITION TO ENSURE THAT IT SHOULD PERFORM AS EXPECTED AND AS REQUIRED.

FROM A SYSTEM DEVELOPMENT PERSPECTIVE, SYSTEM TESTING REFERS TO THE TESTING PERFORMED BY THE DEVELOPMENT TEAM (THE PROGRAMMERS AND OTHER TECHNICIANS) TO ENSURE THAT THE SYSTEM WORKS MODULE BY MODULE ('UNIT TESTING') AND ALSO AS A WHOLE. SYSTEM TESTING SHOULD ENSURE THAT EACH FUNCTION OF THE SYSTEM WORKS AS EXPECTED AND THAT ANY ERRORS (BUGS) ARE NOTED AND ANALYZED. IT SHOULD ADDITIONALLY ENSURE THAT INTERFACE FOR EXPORT AND IMPORT ROUTINES, FUNCTION AS REQUIRED. SYSTEM TESTING DOES NOT CONCERN ITSELF WITH THE FUNCTIONALITY OF THE SYSTEM AND WHETHER THIS IS APPROPRIATE TO MEET THE NEEDS OF THE USERS. HAVING MET THE CRITERIA OF THE TEST PLAN THE SOFTWARE MAY THEN BE PASSED FOR USER ACCEPTANCE TESTING.

THE VARIOUS TESTING METHODOLOGIES PERFORMED FOR THIS SYSTEM IS:

- UNIT TESTING
- Integration Testing
- White Box Testing
- Black Box Testing

Unit testing

In computer programming, a unit test is a procedure used to validate that a particular module of source code is working properly. The idea about unit test is to write test cases for all functions and methods so that whenever a change causes a regression, it can be quickly identified and fixed. Ideally, each test case is separate from the others; constructs such as mock object can assist in separating unit tests. This type of testing is mostly done by the developers and not by end-users.

The goal of unit testing is to isolate each part of the program and show that the individual parts are correct. Unit testing provides a strict, written contract that the piece of code must satisfy. As a result, it affords several benefits. The goal of unit testing is to isolate each part of the program and show that the individual parts are correct. Unit testing provides a strict, written contract that the piece of code must satisfy. As a result, it affords several benefits and allowed to correct the following errors.

1. Mixed mode operations
2. Incorrect initialization
3. Incorrect symbolic representation of the expression
4. Simplified integration
5. Facilitated for the various changes made to the system

IntegrationTesting:

Integration testing can proceed in a number of different ways, which can be broadly characterized as **top down** or **bottom up**. In **top down integration testing** the high level control routines is tested first, possibly with the middle level control structure present only as **stubs**. Subprogram **stubs** are incomplete subprograms which are only present to allow the higher level control routines to be tested.

Top down testing can proceed in a **depth-first** or **breadth-first** manner. For depth-first integration each module is tested in increasing detail, replacing more and more levels of details with actual code rather than stubs. Alternatively breadth-first would proceed by refining all the modules at the same level of the control throughout the application.

In practice a combination of the two techniques would be used. At the initial stage all the modules might be only partly functional, possibly being implemented only to deal with non-erroneous data. These would be tested in breadth-first manner, but over a period of time each would be replaced with successive refinements which were closer to the full functionality. This allows depth-first testing of a module to be performed simultaneously with breadth-first testing of all the modules.

The other major category of integration testing is **bottom up integration testing** where an individual module is tested from a test harness. Once a set of individual modules have been tested they are then combined into a collection of modules, know as **builds**, which are then by a second test harness. This process can continue until the build consists of the entire application.

This second approach is used in this project where the individual modules that are-Mobile Call Status, Mobile Time Retrieval and Internet connectivity are first developed and then later they were integrated into one application and tested for the results.

White Box Testing:

White box testing is testing from the inside—tests that go in and test actual program structure.

Basis path testing: Very simply, test every statement in the program at least once. You'll note that the testing department at FCC chose test cases that did this; the entire execution tree was covered.

Basis path testing is MANDATORY—so much so that there are software products written especially to assist in it.

❖ Profiling: there are a lot of tools—often included with compilers—which show where the CPU is spending most of its time in a program. Naturally, the busiest parts of the program are the ones you want to test most.

❖ Loop tests: Exercise each DO, WHILE, FOR and other repeating statements several times.

❖ Input tests: as the old saying goes—garbage out, garbage out. If a procedure receives the wrong data, it's not going to work. Each procedure should be tested to make certain that the procedure actually received the data you sent to it. This will spot type mismatches, bad pointers, and other such bugs.

Here in this project each decision path is check and all the loops are executed separately to ensure that the program is logically correct and has exited right time

Black Box Testing:

Black box testing, concrete box or functional testing is used in computer programming, software engineering and software testing to check that the outputs of a program, given certain inputs, conform to the functional specification of the program.

The term black box indicates that the internal implementation of the program being executed is not examined by the tester. For this reason black box testing is not normally carried out by the programmer. In most real-world engineering firms, one group does design work while a separate group does the testing.

Boundary value analysis is a technique of black box testing in which input values at the boundaries of the input domain are tested. It has been widely

recognized that input values at the extreme ends of, and just outside of, input domains tend to the cause errors in system functionality.

In boundary value analysis, value at and just beyond boundaries of the input domain are used to generate test cases to ensure proper functionality of the system.

Advantages of Black Box Testing

- More effective on larger units of code than glass box testing
- Testing needs no knowledge of implementation, including specific programming language
- Tester and programmer are independent of each other
- Test are done from a user's point of view
- Will help to expose any ambiguities or inconsistencies in the specifications
- Test cases can be designed as soon as the specifications are complete
  In this project all the function are tested to check whether all of them are working properly. The performance rate is verified by considering response time and speed. Hence the error are identified and corrected.

QUALITY ASSURANCE

Quality assurance comprises all those planned and systematic

actions necessary to provide confidence that a structure, system or component will perform satisfactorily is service.

Quality assurance includes formal view of care, problem definition, corrective actions to remedy any deficiencies and evaluation of actions that to be taken.

The function of software quality that assures that the standards, processes, and procedures are appropriate for the project and are correctly implemented. This is an "umbrella activity" that is applied throughout the engineering process.

Quality software is reasonably bug-free, delivered on time and within budget, meets requirements and/or expectations, and is maintainable.

The system is developed such that it ensures all the level of quality. It checks whether a user friendly environment is provided to the users and that there is a reliable, accurate and efficient flow of data within the system. The system also checks that due it contains the level of security required for the user. Hence as long as there is no hardware complaint, there is no problem with the software.

6.2 SYSTEM IMPLEMENTATION

Plan:

Implementation is the state in the project where the theoretical design is turned into a working system. The most crucial stage in achieving a new successful system and giving confidence on the new system for the users that will work efficiently and effectively. The system is implemented only after thorough testing is done and if it is found to work according to the specification.

It involves careful planning, investigation of the current system and is constraints on implementation, design of methods to achieve changeover, and evaluation of the changeover methods apart from planning. Two major tasks for preparing the implementation are educating, training the users and testing the system.

Implementation plan preparation

The implementation process begins with the preparation of plan for implementation. According to this plan other activities are carried out. In this plan discussion has been made regarding the equipment, resources and how to test the activities. Thus a clear planner prepared for the activities.

Equipment Acquisition

According to the above plan the necessary equipment have to be acquired to implement the new system, which would include all the requirements for installing and maintaining .Net framework, VB.net, SQL server,

Program code preparation

One of the most important development activities is coding or programming. The system flowcharts and other charts are converted into modular programs. They have to be compiled, tested and debugged.

User training and documentation

Once the planning has been completed the major effort in the computer department is that the user department must consist of educated and trained staff as the system becomes more complex. The success of the system depends upon how they are operated and used the system.

Thus the quality of training the personnel is connected to the success of the system. Implementation depends upon the right people being trained at the right time. Education involves creating the right atmosphere and motivating the user. Staff education should encourage the participation of all the staff.

Changeover

Changeover is the change of moving over from the old system to the new computerized system. In order that this is done all the files have to be converted to the new format. The accuracy of the conversion is of utmost importance both to user confidence in the system and to effective operation. When the files have been set up on the computer, the changeover can take place. There are several possible methods of doing this.

E.g. direct changeover, parallel running, pilot running, and staged changeover.

This method is the complete replacement of the old system by new, in one move. When direct changeover is planned, system tests and training should be comprehensive and changeover itself is planned in detail.

Parallel Running:

Parallel running or operation means processing current data by both the old and new systems to cross check the results.

The old system is kept alive and operational until the system has been proved for at least one system cycle, using full live data in the operational environment of place, people, equipment and time. It allows the result of the new system to be compared with the old system before the acceptance by the user. Parallel operation does not allow much time or learning and testing activities.

Staged Changeover:

A staged changeover involves a series of limited size direct changeovers. The new system being introduced piece by piece. A complete start, a logical section is committed to the new system while the remaining parts or sections will be processed by the old system.

In this project, direct changeover is applied where the entire system is implemented directly after it has been developed.

SYSTEM MAINTENANCE:

Maintenance

The term "Software Maintenance" is used to describe software engineering activities. Maintenance activities involve making enhancements to software products, adapting to new environments and correcting problems. Software product enhancements may involve providing new functional capabilities, improving user displays and nodes of interaction, upgrading external documents and internal documentation or upgrading the performance characteristics of a

system. Adaptation of software to a new environment may involve moving the software to a different machine, or for instance, modifying the software to accommodate a new telecommunication protocol or an additional disk drives. Problem correction involves modification and revalidation of software to correct errors.

Many activities performed during software development enhance the maintainability of a software product. They are:-

Analysis activities:

The analysis phase of software development is concerned with determining customer requirements and constraints and establishing feasibility of the product.

- ❖ Develop standards and guidelines
- ❖ Set milestones for supporting documents
- ❖ Specify quality assurance procedures
- ❖ Identify likely product enhancements
- ❖ Determine resources required for maintenance
- ❖ Estimate maintenance costs

Architectural Design Activities:

- ❖ Emphasize clarity and modularity as design criteria
- ❖ Design to ease likely enhancement
- ❖ Use standardized notations to document, data flows, functions, structure and interconnections
- ❖ Observe the principles of information hiding, data abstraction and top-down hierarchical decomposition

Detailed Design Activities

- ❖ Use standardized notations to specify algorithms, data structures and procedure interface specifications
- ❖ Specify side effects and exception handling for each routine

Implementation activitie:s

- ❖ Use single entry, single exit constructs
- ❖ Use standard indentation of constructs
- ❖ Use simple, clear coding style
- ❖ Use symbolic constants to parameterize routines
- ❖ Provide margins on resources
- ❖ Provide standard documentation
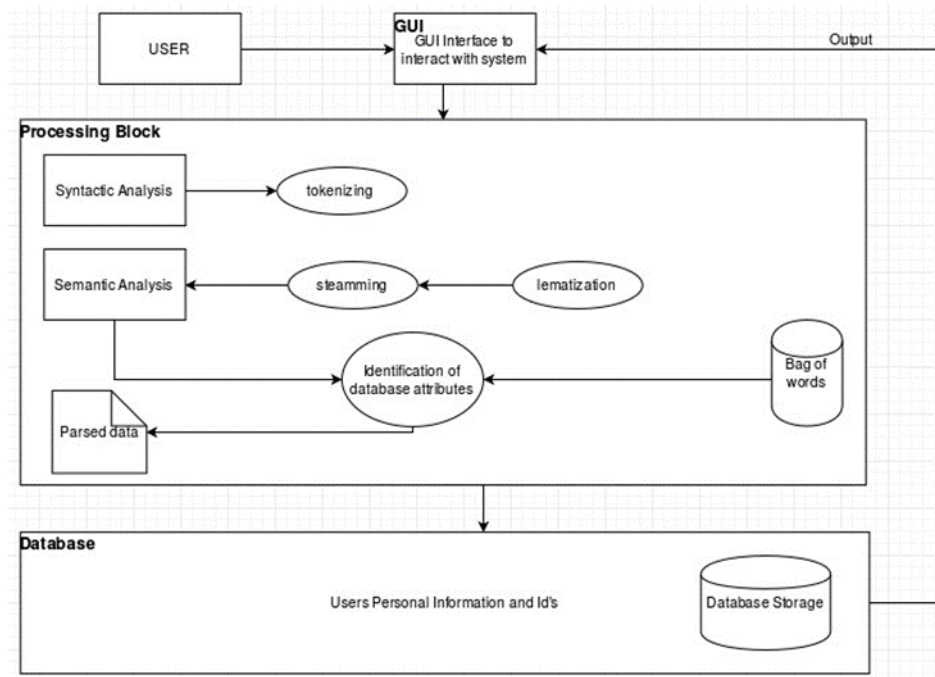- ❖ Follow standard internal commenting guidelines

Other activities:

- ❖ Develop a maintenance guide
- ❖ Develop a test suite
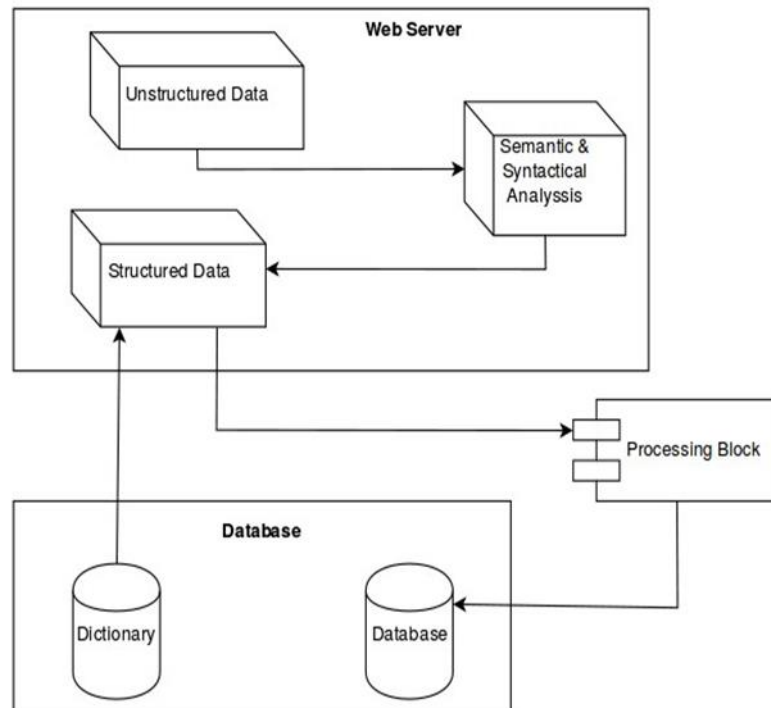- ❖ Provide test suite documentation

CHAPTER 7

SYSTEM DESIGN

7.1 DATA FLOW DIAGRAM (BLOCK DIAGRAM)

This diagram shows what happens inside a computer when user interacts with the system to show database.

7.2 DATABASE DESIGN

This diagram shows the design of database in Web Server.

7.3 INPUT DESIGN

Input design is the process of converting user-originated inputs to a computer-based format. Input design is one of the most expensive phases of the operation of computerized system and is often the major problem of a system.

In the project, the input design is made in various window forms with various methods.

- STUDENT DETAIL
- MARKS DETAILS

7.4 OUTPUT DESIGN

Output design generally refers to the results and information that are generated by the system for many end-users; output is the main reason for developing the system and the basis on which they evaluate the usefulness of the application. In any system, the output design determines the input to be given to the application.

In the project, the output design is made in various window forms with various methods.

- View Graph
- RESULTS

CHAPTER 8
CONCLUSION

8.1 BIBLIOGRAPHY

Reference:

1.Jessica Simko , "How Hiring Managers Make Decisions" http://www.careerealism.com/hiring-managers-decisions/ Vinayak Joglekar , "Ranking Resumes using MachineLearning"
2.https://vinayakjoglekar.wordpress.com/2014/06/24/ranking-resumes-using-machine Peter Gold , "Artificial Intelligence Recruiting" https://www.linkedin.com/pulse/artificial-intelligence-recruiting-peter-gold
Turbo Ricruit ,"Automated Application Processing","Better candidate experience","Matching Job Descriptions to Resumes"
3.http://www.turborecruit.com.au/benefits-of-artificial-inteligence-for-recruitment/

[4]   Rchillies, http://www.rchillies.com

[5]   Belong.co, http://www.belong.co

[6]   ALEX System , http://www.hireability.com/alex/

[7]   Turbo Ricruit , http://www.turborecruit.com.au/intelligent-searching/

[8]   Revolvy , http://www.revolvy.com/main/index.php?s=Parse%20tree

[9]   Student Thesis "Information Quality Management inInformation Extraction:A
Survey"..http://www.rn.inf.tudresden.de/uploads/Studentische_Arbeiten/Belegar beit_Janse n_Nicolas.pdf

8.2 APPENDIX

A) SCREENSHOTS

Categorywise Distribution
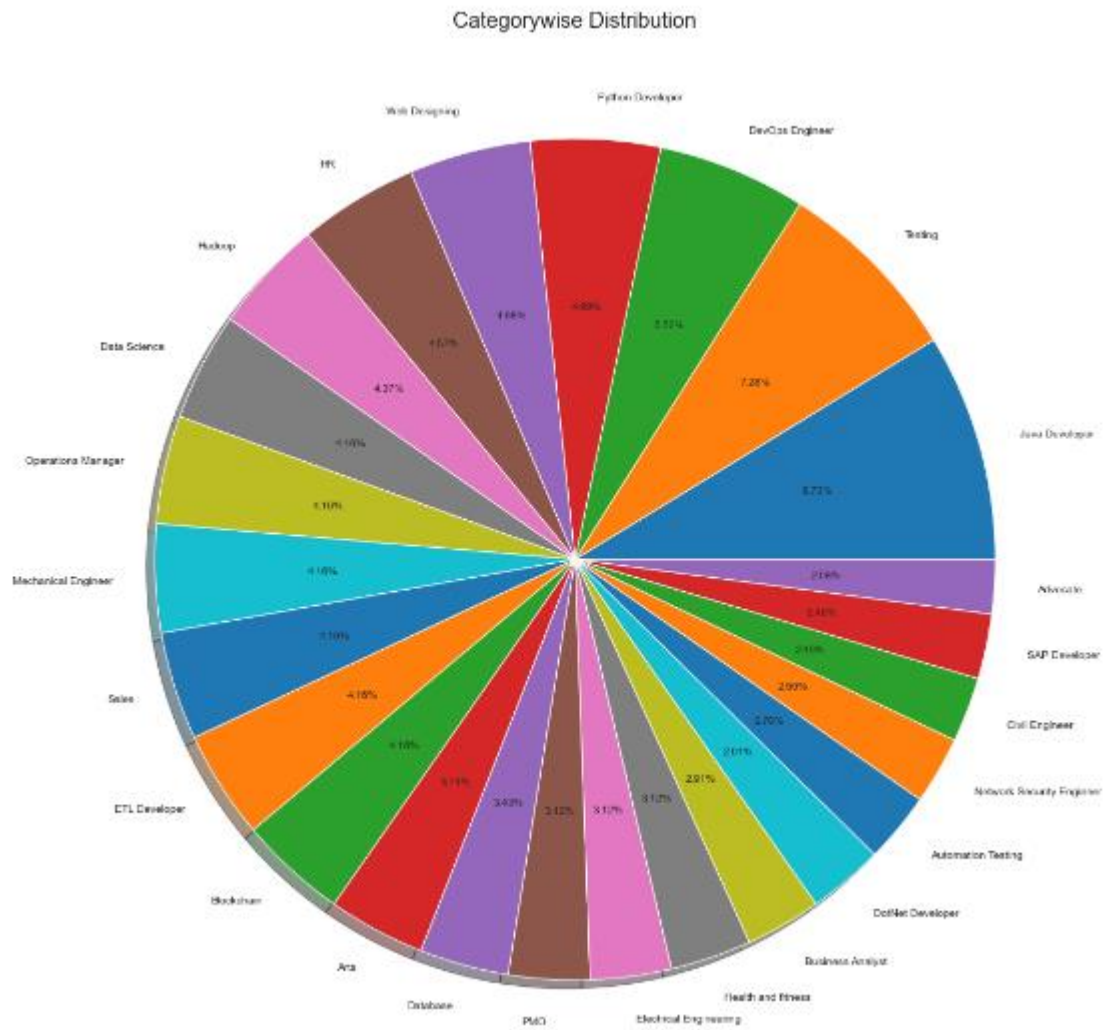
B)REPORT
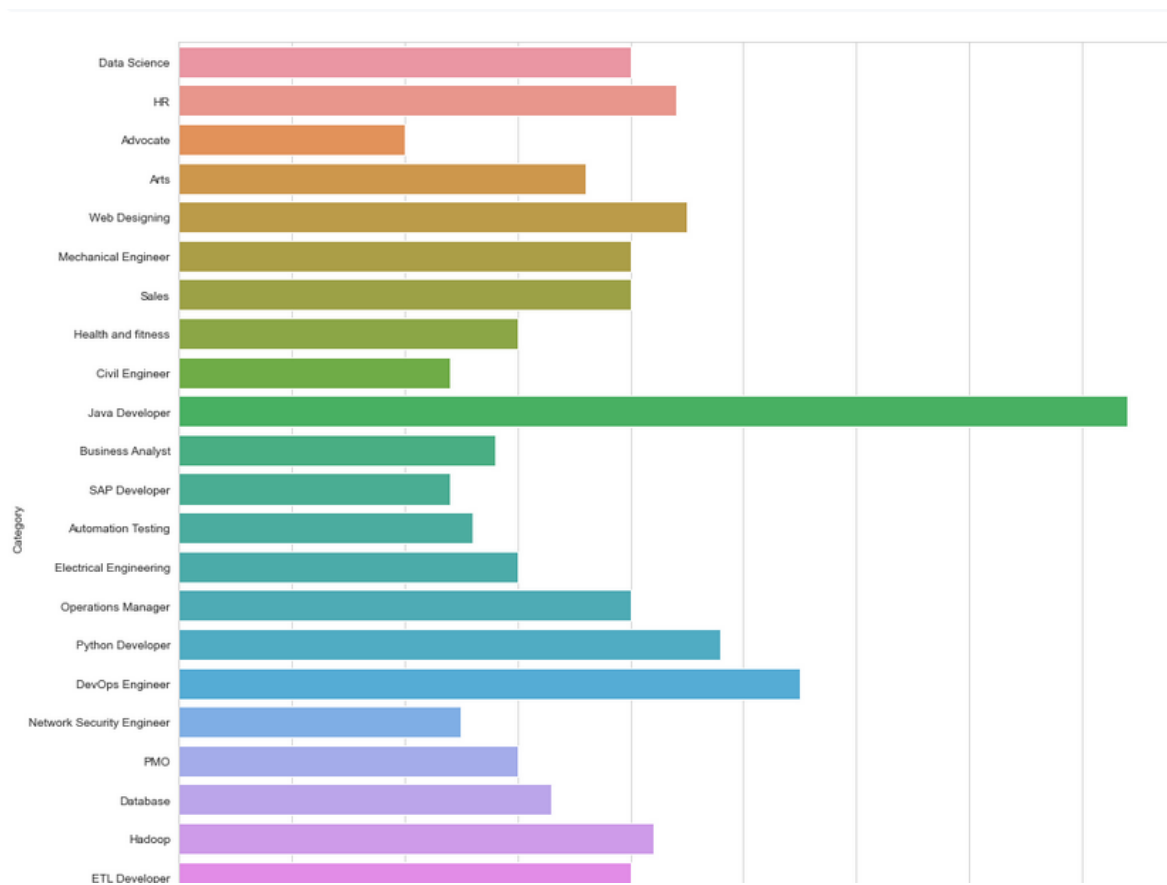
C)SAMPLE CODE

```
import os
import zipfile
import pandas as pd
import seaborn as sns
import numpy as np
os.chdir("/home/linux/Downloads")
from sklearn import tree
import matplotlib.pyplot as plt
from sklearn.tree import DecisionTreeClassifier
from sklearn.datasets import make_multilabel_classification
```

```python
from sklearn.multioutput import MultiOutputClassifier
with zipfile.ZipFile("archive(3).zip") as z:
    with z.open("ibm.csv") as f:
        df=pd.read_csv(f)
df.head(50)
cols=df.columns[:35]
colours = ['#000099', '#ffff00']# specfiy color yellow missing,blue is not
missing
sns.heatmap(df[cols].isnull(), cmap=sns.color_palette(colours))
for col in df.columns:
    pct_missing = np.mean(df[col].isnull())
    print('{} - {}%'.format(col, round(pct_missing*100)))
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(df[["Age","JobLevel","WorkLifeB
alance","YearsWithCurrManager"]],df[["PercentSalaryHike"]],test_size=0.35,ra
ndom_state=42)
from sklearn.preprocessing import StandardScaler
scale = StandardScaler()
x=df[["Age","JobLevel"]]
y=df[["PercentSalaryHike"]]
scaledx = scale.fit_transform(x)
clf
=DecisionTreeClassifier(criterion='gini',splitter='best',max_depth=10,random_s
tate=None)
clf.fit(scaledx,y)
import category_encoders as ce
encoder
=ce.OrdinalEncoder(cols=["Age","JobLevel","WorkLifeBalance","YearsWithC
urrManager"])
```

```python
x_train = encoder.fit_transform(x_train)
x_test = encoder.transform(x_test)
clf =DecisionTreeClassifier(criterion='gini',splitter='best',max_depth=10,random_state=None)
clf.fit(x_train,y_train)
y_pred=clf.predict(x_test)
from sklearn.metrics import accuracy_score
accuracy_score(y_pred,y_test)*100
clf_en = DecisionTreeClassifier(criterion='entropy', max_depth=3, random_state=0)
clf_en.fit(x_train,y_train)
y_pred_en=clf_en.predict(x_test)
accuracy_score(y_pred_en,y_test)*100
y_predict=clf.predict(x_train)
accuracy_score(y_predict,y_train)*100
clf.score(x_train,y_train)*100
clf.score(x_test,y_test)*100
plt.figure(figsize=(12,12))
tree.plot_tree(clf.fit(x_train,y_train));
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred_en)
print('Confusion matrix\n\n', cm)
from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred_en))
```