

High-Level Design (HLD)

Salary Prediction

Revision Number – 1.2

Last Date of Revision: 26– 11 -2022

Nitesh Sharma

Document Version Control

Date	Version	Description	Author
26– 11- 2022	1.0	Abstract Introduction Problem Statement	Nitesh
26– 11- 2022	1.1	Design Flow	Nitesh
26 – 11 - 2022	1.2	Performance Evaluation Conclusion	Nitesh

Contents

Document Version Control...	2
Abstract...	4
1. Introduction	5
1.1 Why this HLD Document ?.....	5
1.2 Scope.....	5
2. General Description	6
2.1 Problem Statement.....	6
2.2 Proposed Solution.....	6
2.3 Further Improvements.....	6
2.4 Technical Requirements.....	7
2.5 Data Requirements.....	7
2.6 Tools Used.....	7
2.7 Constraints.....	7
2.8 Assumptions	8
3. Design Flow	8
3.1 Model Creation and Evaluation.....	8
3.2 Deployment Process	8
3.3 Logging	9
3.4 Error Handling.....	9
4. Performance Evaluation	9
4.1 Reusability.....	9
4.2 Application Compatibility.....	9

4.3 Resource Utilization	9
4.4 Deployment	9
5. Conclusion.....	9

Abstract

The prominent inequality of wealth and income is a huge concern especially in the United States. The likelihood of diminishing poverty is one valid reason to reduce the world's surging level of economic inequality. The principle of universal moral equality ensures sustainable development and improve the economic stability of a nation. Governments in different countries have been trying their best to address this problem and provide an optimal solution. This study aims to show the usage of machine learning and data mining techniques in providing a solution to the income equality problem.

Classification has been done to predict whether a person's yearly income in US falls in the income category of either greater than 50K Dollars or less equal to 50K Dollars category based on a certain set of attributes. The Gradient Boosting Classifier Model was deployed which clocked the highest accuracy of 80 %, eventually breaking the benchmark accuracy of existing works.

1. Introduction

1.1 Why this HLD Document ?

The main purpose of this HLD document is to feature the required details of the project and supply the outline of the Model Creation, Evaluation and Deployment. This additionally provides a careful description on how the complete project has been designed end-to-end.

The HLD will :

- Present of the design aspects and define them in detail.
- Describe the user interface being implemented.
- Describe the hardware and software interfaces.
- Describe the performance requirements.
- Include design features and architectural design of the project.
- List and describe the non - functional attributes like :
 1. Security
 2. Reliability
 3. Maintainability
 4. Portability
 5. Reusability
 6. Resource
 7. Utilization

1.2 Scope

The HLD documentation presents the structure of the system, such architectural design, application flow and technology architecture. The HLD uses non-technical terms to technical terms that can be understandable to the administrator of the system.

1.3 Definitions

Term	Description
SP	Salary Prediction

Jupyter - Notebook	It is an interactive computational environment, in which you can combine code execution, rich text, mathematics, plots and rich media.
AWS beanstalk	AWS is a platform as a service (PaaS) that enables developers to build, run, and operate applications entirely in the cloud.

2. General Description

2.1 Problem Statement

The Goal is to predict whether a person has an income of more than 50K a year or not. This is basically a binary classification problem where a person is classified into the >50K group or <=50K group.

2.2 Proposed Solution

To solve the problem, we have created a User interface for taking the input from the user to predict the Salary using our trained ML model after processing the input and at last the output (predicted value) from the model is communicated to the User.

2.3 Further Improvements

We also analyze the data used for training the ML model by considering different angles of business. If we use such information and predict the salary it will help the organization .

2.4 Technical Requirements

As technical requirements, we don't need any specialized hardware for virtualization of the application. The user should have the device that has the access to the web and the fundamental understanding of providing the input.

2.5 Data Requirements

is accessible on the Kaggle within the file format of (.csv). Because the main theme of the project is to induce the expertise of real time issues .

2.6 Tools Used

- Python 3.9 is employed because of the programming language and frameworks like NumPy, Pandas, Scikit - learn and alternative modules for building the model.
- Jupyter - Notebook is employed as an IDE.
- For Data visualizations, seaborn and components of matplotlib are getting used.
- For information assortment prophetess info is getting used.
- Front end development i use streamlit framewok.
- GitHub is employed for version management.
- AWS beanstalk is employed for deployment.

2.7 Constraints

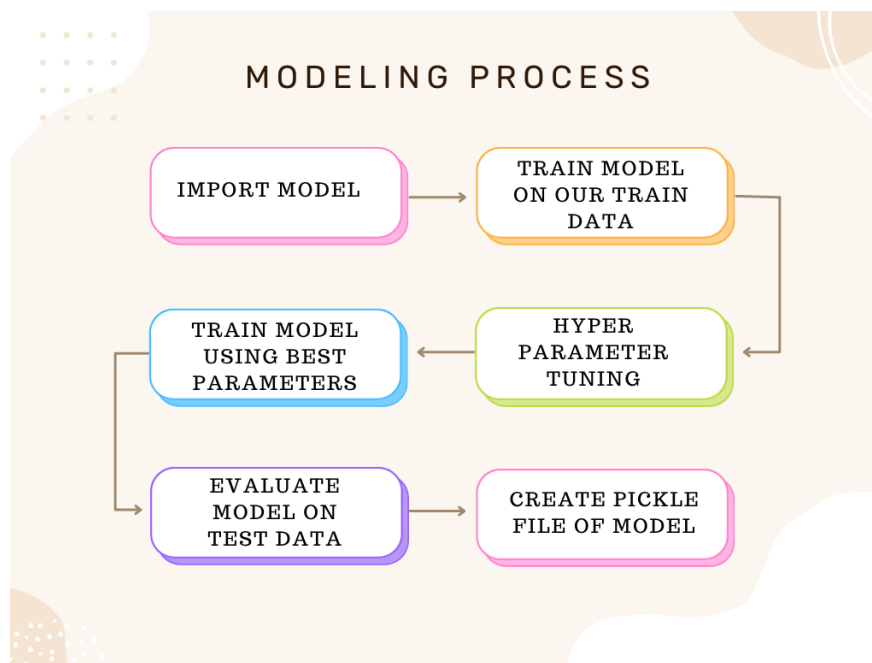
The Salary prediction answer should be user friendly, as automatic as attainable and also the user should not be needed to understand any of the operations.

2.8 Assumptions

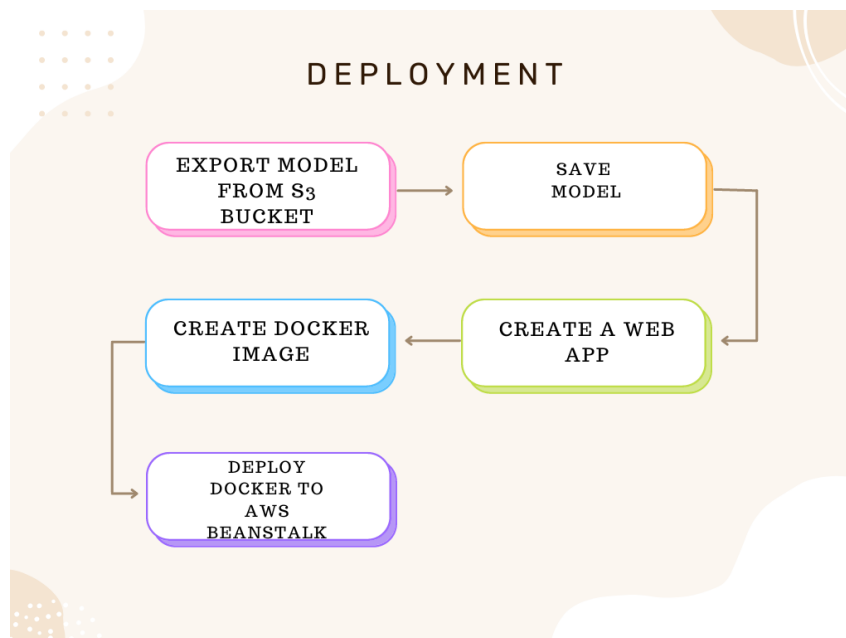
The main objective of the project is to implement the use cases for the new dataset that provides the user the ability to predict Salary. Machine learning model is employed for process the user input for prediction. It is additionally assumed that each one of the aspects of this project has the flexibility to figure along within the approach the designer is expecting.

3. Design Flow

3.1 Modelling Process



3.2 Deployment Process



3.3 Logging

In logging, each time an error or an exception occurs, the event is logged into the system log file with reason and timestamp. This helps the developer to debug the system bugs and rectify the error.

3.4 Error Handling

Once the error occurs, the reason is logged into the log file with timestamp to rectify and handle it.

4. Performance Evaluation

4.1 Reusability

The code written and the components used should have the ability to be reused with no problems.

4.2 Application Compatibility

The different parts of the system are communicating or using Python as an interface between them. All the components have its own tasks to perform and it is the job of a Python to ensure proper transfer of data.

4.3 Resource Utilization

When a task is performed, it'll doubtless use all the process power offered till the process is finished.

4.4 Deployment

The model can be deployed using any cloud services such as Microsoft Azure, Amazon web services, Heroku, Google cloud, etc.

5. Conclusion

The salary prediction system is help the organization and government in taking decision based on people's salary based on changes in their occupation and education etc.