**Salary Predictor**

**Project Synopsis**

**Industrial Training**

**BACHELOR OF COMPUTER APPLICATION**

|  |  |
| --- | --- |
| **PROJECT GUIDE:** | **SUBMITTED BY:** |
| **Mrs. Rohaila Naaz** | **Ayush jain(TCA1901063)** |
|  |  |
|  |  |



**FACULTY OF ENGINEERING & COMPUTING SCIENCES**

**TEERTHANKER MAHAVEER UNIVERSITY, MORADABAD**



|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Table of Contents** |  |
| **1** | **Project Title ...........................................................................................................................................** | | **3** |
| **2** | **Domain ..................................................................................................................................................** | | **3** |
| **3** | **Problem Statement ...............................................................................................................................** | | **3** |
| **4** | **Project Description ................................................................................................................................** | | **3** |
| **4.1** | | **Scope of the Work ........................................................................................................................** | **3** |
| **4.2** | | **Project Modules ............................................................................................................................** | **3** |
| **5** | **Implementation Methodology ..............................................................................................................** | | **3** |
| **6** | **Technologies to be used .......................................................................................................................** | | **4** |
| **6.1** | | **Software Platform .........................................................................................................................** | **4** |
| **6.2** | | **Hardware Platform .......................................................................................................................** | **4** |
| **6.3** | | **Tools ..............................................................................................................................................** | **4** |
| **7** | **Advantages of this Project ....................................................................................................................** | | **4** |
| **8** | **Future Scope and further enhancement of the Project .......................................................................** | | **4** |
| **9** |  | **Conclusion .........................................................................................................................................** | **5** |
| **10** |  | **References ........................................................................................................................................** | **5** |





1. **Project Title**

Salary Predictor

1. **Domain**

The domain of this project is Machine learning where we use Sklearn library under which we have use linear model and model selection for performing linear regression , and train test .

1. **Problem Statement**

Machine learning is the subset of artificial intelligence ,using machine learning we build algorithm or models which can predict future output and results based on past experience and trained dataset. By different output and training dataset our algorithm improves their learning with time. The project problem statement is to find or predict the salary for the employees. This model will help normal peoples to predict their salaries according to their experience .

1. **Project Description**

The objective of this project to learn implementation of any machine learning model and understand the workflow of any model. This project is small contribution toward the community.

* 1. **Scope of the Work**

User can predict the salary by just giving required input. The model can’t predict output until all required input is filled. Filling wrong data or NAN data leads to error or wrong output.

* 1. **Project Modules**

1. **Data preparation and pre-processing** : In this module we fetch dataset and pre-process dataset by cleaning and removing the unrequired data.
2. **Data splitting :** In this module we split our dataset into 2 part ,one will train dataset which will use to train our model and second one is test dataset which will use score or test our model.
3. **Data visualization:** In this module we visualize dataset using matplotlib and seaborn.

1. **Modelling :**  In this module we will set our dataset in machine learning algorithm, predict output, score our model and teach model to improve on the biases of different output score.
2. **Model deployment : I**n this module we will deploy our model using Flask on web.
3. **Implementation Methodology**

The implementation of this project is divided into 7 parts:

1. **Data preprocessing :** The first step is to import the dataset into our IDE, then filter the dataset by removing the unrequired data, duplicate row or data set and removing NaN data. Then filtered dataset is divided into subparts Train dataset and Test dataset.
2. **Choose the right estimator/algorithm for our problems:** After getting our data ready we choose preferred machine learning algorithm for problems and estimator.
3. **Fit the Model :** After choosing the right machine learning algorithm, we will fit our train dataset into the model and use it to make prediction on our data.
4. **Evaluating a Model**: After fitting data into model. We will be evaluating our model; it helps to find best machine learning model that represent our dataset and how well the chosen model work in future.
5. **Improve the model**: By putting different estimator we will score our model and try to improve the model on the biases of different output. At the end we will put all together.
6. **Deployment:**  This is the last step of our project where we deploy our model on web application using different framework but, in this project, we used flask for deployment.



**DATAFLOW DIAGRAMS**

The entire working or the flow of the data can be divided into two groups for better understanding They are 1 DFD-L0, 2. DFD-L1

Send the data

Enter details

Salary Prediction

Server

User

**DFD-L0**

This is the initial idea for the data. The data has to be flown from user to server and form server to the user for the prediction of the Salary.

Input details

**Server**

Feed the values

**User**

Match values

With dataset

Predict the

salary

Send the details

**DFD-L1**

****

This is the process or the idea where the data has been used to predict the salary by following several steps like registration (for new users), Feed the values(entering and storing values), Server(to store them), match the values(Finding probability) and finally predict the salary (Final result).



**6** **Technologies to be used**

**6.1** **Software Platform**

1. **Front-end**

Flask

1. **Back-end**

Mysql

1. **Framework**

Flask

**6.2** **Hardware Platform**

**RAM:**4GB

**Hard Disk:** 250Gb

**OS: windows, Linux, Mac etc**

**6.3** **Tools**

Python 3.2, Anaconda, Jupyter lab

Numpy , Pandas

**Machine Learning**

Linear Regression

Sklearn

Mean absolute error for finding error

Matplotlib

Seaborn

**7** **Advantages of this Project**

* User can predict the salary of an employee by just giving some required input experience data according to their requirement
* User can access the model at time and from anywhere.
* There is not any prerequisite to use this project and the UI quite simple to use.
* It is free of cost

**8** **Future Scope and further enhancement of the Project**

The future of this project is, we can make a build app for this project and automate the project by automatically putting the input at different time to predict output. With the increase of different dataset, the accuracy of output will automatically increase.

**9 Conclusion**

By building this project I learn machine learning and different tools for machine learning development. I also understand the workflow of machine learning algorithm and how to deploy any model. Also this project helps me to enhance my resume and it will my small contribution towards community.

**10 References**



1. https://trainings.internshala.com/progress/home/machine-learning
2. [https://www.kaggle.com/ronitf/salary-experience-uci](https://www.kaggle.com/ronitf/heart-disease-uci)