

**Department of Computer Science**

**and Engineering- Data Science**

ANNEXURE -VII

INTERSHIP REPORT FORMAT

STIUDENT INTERSHIP PROGRAM (SIP) REPORT

**PROJECT**

Submitted by

Name of student = Payal Shrikant Bajantri

PRN = 202401120064



**CERTIFICATE**

**This is to certify that the “student internship program (SIP)” report submitted by PAYAL SHRIKANT BAJANTRI. PRN 202401120064 is work done by her and is submitted during 2025-26 academic year .**

**Faculty Mentor**

**School - Internship Coordinator**

**School :HOD**



Table of Contents

|  |  |  |
| --- | --- | --- |
|  | CONTENT | PAGE NO |
|  | Abstract |  |
| 1 | Introduction |  |
| 2 | Methodology |  |
| 2.1 | Data Analysis and Preparation |  |
| 2.2 | Model Training |  |
| 2.3 | Application Development |  |
| 3 | Results |  |
| 4 | Conclusion and Future Scope |  |



. Abstract

This report details the development of a machine learning model to predict employee salaries based on years of experience and job rating. Using a linear regression approach, the model was trained on a dataset to establish a relationship between these features and salary. The final model is integrated into a user-friendly web application built with Streamlit, which allows users to input data and receive real-time salary predictions. This project demonstrates the practical application of machine learning for business and human resource .



1. INTRODUCTION

In modern business, data-driven decision-making is crucial. Accurately predicting employee salaries is a key task for companies to ensure fair compensation, manage budgets, and attract talent. This project aims to address this need by creating a predictive model using machine learning. The model's purpose is to provide a reliable salary estimate based on two key features: the number of years of experience and a job performance rating. The project culminates in a simple, intuitive web application for practical use

1. Methodology

The development process was divided into three main phases: data preparation, model training, and application deployment.

* 1. Data Analysis and Preparation

The dataset for this project was not explicitly provided in the user's prompt, but for a linear regression model, it would typically contain three columns: "Years of Experience," "Job Rate," and "Salary." The data was first analyzed to understand the relationship between the features and the target variable. This analysis involved checking for correlations between the variables and visualizing the data to identify any trends. The data was then cleaned to handle any missing values and outliers. The features (years and job rate) and the target variable (salary) were separated and prepared for the training phase.

**2.2 Model Training**

A linear regression model was chosen for its simplicity and interpretability, as it is well-suited for predicting a continuous numerical value like salary. The model was trained using the scikit-learn library in Python. It learns a linear relationship, represented by the equation: Salary=(β0​)+(β1​∗Years)+(β2​∗JobRate)+

where β0​ is the intercept, β1​ and β2​ are the coefficients for each feature, and ϵ is the error term. The trained model was then saved as a serialized file (linearmodel.pkl) using joblib for later use in the application .

* 1. Application Development

The user interface was built using the Streamlit framework, which is excellent for creating data applications with minimal code. The application does the following:

* Initializes the app title and a brief description.
* Creates two number input widgets for the user to enter years of experience and jobrate.
* Loads the pre-trained linear regression model from the linearmodel.pkl file.
* Includes a "Predict" button to trigger the prediction.
* When the button is pressed, the user inputs are converted into a NumPy array, which is the required format for the model's predict() method.
* The model generates a salary prediction, which is then displayed to the user.

4.Conclusion and Future Scope

This project successfully demonstrates the end-to-end process of building a machine learning-powered application, from model training to deployment. The Streamlit application provides a simple and effective tool for salary estimation.

For future improvements, the model could be enhanced by:

* Incorporating more features such as department, location, or education level.
* Experimenting with more complex models like Random Forest or Gradient Boosting to potentially improve accuracy.
* Gathering a larger, more diverse dataset for training to make the model more robust.
* Adding data visualization to show how years of experience and job rate influence the the predicted salary.

[**GitHub - payalbajantri/internship**](https://github.com/payalbajantri/internship)