SALARY PREDICTION USING LINEAR REGRESSION THROUGH HARDWARE DISCRIPTION LANGUAGE

A mini project report submitted in partial fulfilment of the requirements for the degree of

SRIKAKULAM – 532410

March – 2024

Bachelor of Technology

In

Electronics and Communication Engineering

By

S190607 - A. Vishnu suji Priya

S190547 - S.Gayatri
S190374 - V. Roshini

S190570 - P.charan

S191105 - P.Taraka Ratna

S190022 - P. Abhinay ARYA

Under the supervision of

M. Ramu M.Tech

Electronics communications Engineering

Department Of Electronics and Communication Engineering

Rajiv Gandhi University of Knowledge Technologies

Cetificate

This is to certify that the work entitled " **SALARY PREDICTION USING LINEAR REGRESSION THROUGH HARDWARE DISCRIPTION LANGUAGE"** bonafide record of authentic work carried out by S190607, S190547, S190374, S190570, S191105, S190022 under my supervision and guidance for the partial fulfilment of the requirement of the award of the degree of Bachelor of Technology in the department of Electronics and Communication Engineering

The results embodied in this work have not been submitted to any other university or institute for the award of any degree or diploma. This thesis, in our opinion, is worthy of consideration for the award of the degree of Bachelor of Technology in accordance with the regulations of the institute.

M.Ramu, M.Tech ECE
Assistant Professor,
Department of ECE,
RGUKT-Srikakulam.

Mr. N. Ramesh Babu,
M.Tech Head of the
Department, Department of
ECE,
RGUKT-Srikakulam

Acknowledgement

I would like to express my special thanks of gratitude to my project guide **M. Ramu** Assistant Professor, Department of ECE, who gave me the golden opportunity to do this wonderful project on the topic

" SALARY PREDICTION USING LINEAR REGRESSION THROUGH HARDWARE DISCRIPTION LANGUAGE" , which also helped me in doing a lot of research and I came to know about so many new things I am really thankful to them.

Electronics and Communication Engineering

IIIT-SRIKAKULAM

Date: 14-06-2024

Place : - NUZVID

TABLEOFCONTENTS

Abstract

Salary prediction using linear regression remains a pivotal tool in labor economics and human resource management, bridging historical foundations with modern applications. This review traces the development from the classic human capital earnings function to current machine learning adaptations, underscoring linear regression's enduring relevance due to its simplicity and clarity. By examining key methodologies and the integration of regularization techniques, we highlight how these models are employed to predict salaries, support HR decisions, and guide economic research. Despite the advent of more sophisticated algorithms, linear regression continues to be invaluable for its interpretability and practical utility in addressing challenges such as data quality and evolving market dynamics. Future directions emphasize the need for continuous model updating and ethical considerations to ensure fair and accurate salary predictions.

Software tools: Jupyter Notebook

Libraries Used: Matplotlib, Open CV, Numpy

1. INTRODUCTION

Salary prediction is a vital component in understanding labor market dynamics, playing a crucial role in career planning, employee retention, and compensation strategies. It enables individuals to make informed career decisions and helps organizations design competitive salary packages that attract and retain talent. Among the various predictive techniques, linear regression stands out for its simplicity and ease of interpretation. Rooted in the foundational human capital earnings function, linear regression models have evolved to incorporate modern machine learning enhancements, maintaining their relevance and utility in accurately forecasting salaries in a rapidly changing job market.

1.1. Scope

The scope of this project encompasses the application and evaluation of linear regression techniques in predicting salaries within diverse economic and occupational contexts. We aim to explore the utility of linear regression, from its traditional use in human capital earnings models to its integration with advanced regularization methods like Ridge, Lasso, and Elastic Net. This project will analyze how these models perform in predicting salaries based on various factors such as education, experience, skills, and geographic location. Additionally, we will assess the models' applicability in practical domains including human resource management, career counseling, and economic research. The project will also address challenges related to data quality and market dynamics, providing recommendations for continuous model updating and the ethical use of salary prediction models. Through comprehensive analysis and comparison, this project seeks to reaffirm the significance of linear regression in salary prediction while highlighting future directions for enhancing its accuracy and fairness.

1.2. Project Requirements

To successfully execute the project on salary prediction using linear regression, several key requirements must be met. These requirements can be categorized into data, software, and technical components:

Data Requirements

1. **Dataset Collection**:
    - **Salary Data**: Obtain comprehensive datasets containing salary information along with attributes such as education level, years of experience, skills, job titles, and geographic location.
    - **Demographic Data**: Include demographic information if available (e.g., age, gender) to analyze disparities and ensure equitable predictions.
2. **Data Quality**:
    - Ensure that the datasets are clean, with minimal missing values, and accurately represent the target population.
    - Perform data preprocessing tasks such as handling missing values, encoding categorical variables, and normalizing numerical features.
3. **Data Sources**:
    - Identify reliable sources for salary and related data, such as public labor market databases, company payroll records, or reputable salary survey reports.
    - Secure necessary permissions for using proprietary or sensitive data, ensuring compliance with data privacy and ethical guidelines.

Software Requirements

1. **Programming Languages and Tools**:
    - **Python**: Utilize Python for its extensive libraries and frameworks suited for data analysis and machine learning (e.g., pandas, scikit-learn, TensorFlow/Keras).
    - **Jupyter Notebook**: Use Jupyter Notebooks for interactive data exploration and model development.
2. **Machine Learning Libraries**:
    - **Scikit-learn**: Implement linear regression models and regularization techniques using Scikit-learn’s robust machine learning tools.
    - **NumPy and Pandas**: Employ these libraries for efficient data manipulation and numerical computations.
    - **Matplotlib and Seaborn**: Use these libraries for data visualization to illustrate model performance and insights.
3. **Integrated Development Environment (IDE)**:

- Utilize an IDE such as PyCharm or Visual Studio Code for efficient code development and debugging.

Technical Requirements

1. **Model Development and Validation**:
   - Develop and evaluate linear regression models, including basic, multiple, and regularized regression techniques.
   - Split the dataset into training and testing sets to validate the models and assess their predictive performance.
2. **Performance Metrics**:
   - Define and compute key performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) to evaluate the accuracy and robustness of the models.
3. **Computational Resources**:
   - Ensure access to sufficient computational resources, including a computer with adequate processing power and memory to handle large datasets and perform intensive computations.
4. **Documentation and Reporting**:
   - Maintain detailed documentation of the project's development process, including data sources, preprocessing steps, model configurations, and performance results.
   - Prepare a comprehensive report and presentation summarizing the findings, methodologies, and practical implications of the project.

Optional Requirements

1. **Advanced Techniques**:
   - Explore the integration of advanced machine learning techniques such as ensemble methods or neural networks to compare with linear regression models.
   - Consider implementing automated machine learning (AutoML) tools for optimizing model selection and hyperparameters.
2. **Ethical and Fairness Considerations**:
   - Conduct fairness assessments to ensure that salary predictions are unbiased and equitable across different demographic groups.

- Develop guidelines for the ethical use of salary prediction models in real-world applications, highlighting the importance of transparency and fairness.

2. PROBLEM DEFINITION

2.1 Existing Method

This existing method using linear regression and its regularized variants provides a structured approach to salary prediction. It highlights the process from data preparation through model training and evaluation, offering a clear pathway to understand how different factors influence salaries. The method's adaptability and interpretability make it a reliable tool for both academic research and practical applications in various domains. For salary prediction, linear regression has long been a favoured method due to its simplicity, interpretability, and effectiveness in capturing relationships between salary and its predictors. Here, we outline an existing method using linear regression, specifically focusing on the traditional human capital earnings function and its extensions with regularization techniques.

Some disadvantages in the existing method:-

1) **Assumption of Linearity:** Linear regression assumes a linear relationship between predictors (like education, experience) and the response variable (salary). If this assumption is not met, the model may provide biased predictions.
2) **Limited Flexibility:** Linear regression models have limited flexibility in capturing complex, non-linear relationships between predictors and salary. This can lead to underperformance when the true relationship is non-linear.
3) **Susceptibility to Outliers:** Linear regression can be sensitive to outliers in the data, which can disproportionately influence the model's coefficients and predictions.
4) **Multicollinearity:** When predictors are highly correlated with each other (multicollinearity), linear regression coefficients can become unstable and difficult to interpret correctly.
5) **Limited Feature Selection:** Linear regression requires careful selection and preprocessing of features. It may not effectively handle large numbers of predictors without feature engineering or dimensionality reduction techniques.

2.2. Proposed Method

Linear regression remains a valuable tool for salary prediction due to its simplicity, interpretability, and effectiveness in capturing linear relationships between predictors and salary. However, to enhance its robustness and address inherent limitations, a hybrid approach integrating modern techniques can be adopted:

1. **Feature Engineering and Transformation:**
   - **Polynomial Features:** Extend linear regression by incorporating polynomial features to capture non-linear relationships that may exist between predictors and salary.
   - **Feature Selection:** Use advanced feature selection methods (e.g., Lasso regularization) to identify and include only the most relevant predictors, mitigating the impact of multicollinearity and improving model interpretability.
2. **Handling Non-linearity and Complex Relationships:**
   - **Ensemble Methods:** Implement ensemble techniques such as Random Forest or Gradient Boosting to capture complex interactions and non-linearities that linear regression may miss.
   - **Neural Networks:** Utilize neural networks for automatic feature extraction and learning hierarchical representations of data, especially when dealing with unstructured or high-dimensional data.
3. **Addressing Data Challenges:**
   - **Outlier Detection and Treatment:** Apply robust statistical techniques to detect and handle outliers that could distort linear regression predictions.
   - **Data Imputation and Cleaning:** Employ rigorous data preprocessing methods to handle missing values and ensure data quality, crucial for accurate salary predictions.
4. **Model Evaluation and Validation:**
   - **Cross-Validation:** Use cross-validation techniques to assess model performance across different subsets of data, ensuring generalizability and reliability of predictions.
   - **Ethical Considerations:** Integrate ethical considerations in data collection and model development to mitigate biases and ensure fairness in salary predictions across demographic groups.
5. **Continuous Model Refinement:**
   - **Iterative Model Improvement:** Continuously refine models based on feedback from stakeholders and updates in data availability, incorporating new predictors and refining existing features.

- **Monitoring and Adaptation:** Implement mechanisms to monitor model performance over time and adapt to changing trends in the job market and economy.

By integrating these approaches, the proposed method enhances the foundational strengths of linear regression while leveraging advancements in predictive modeling to improve accuracy, robustness, and applicability of salary prediction models. This approach not only acknowledges the enduring relevance of linear regression but also embraces innovation to meet evolving challenges in salary prediction.

3 SYSTEM ANALYSIS

Linear regression remains a valuable tool for salary prediction due to its simplicity, interpretability, and foundational role in predictive modeling. To maximize its effectiveness and address potential limitations, the following methodology can be employed:

1. **Feature Engineering and Transformation:**
   - **Polynomial Features:** Extend linear regression by including polynomial features (e.g., squared terms) to capture non-linear relationships between predictors (education, experience, etc.) and salary.
     - **Benefit:** This allows the model to better fit data with non-linear patterns, improving predictive accuracy without departing from the interpretability of linear regression.
2. **Regularization Techniques:**
   - **Ridge Regression:** Incorporate Ridge regression to penalize large coefficients and reduce overfitting, especially when dealing with multicollinearity among predictors.
   - **Lasso Regression:** Utilize Lasso regression to perform feature selection and automatically select the most relevant predictors by shrinking less important coefficients to zero.
     - **Benefit:** Regularization techniques enhance the model's robustness against noise and improve generalizability by preventing overfitting.
3. **Model Evaluation and Validation:**
   - **Cross-Validation:** Implement k-fold cross-validation to assess model performance across different subsets of data.
     - **Benefit:** This technique helps in understanding how well the model will generalize to new data, ensuring reliable predictions in various scenarios.

4. **Handling Data Challenges:**
    - **Outlier Detection and Treatment:** Use robust statistical methods to identify and handle outliers that could skew predictions.
    - **Data Preprocessing:** Clean and preprocess data to handle missing values and ensure data quality before applying linear regression.
        - **Benefit:** Clean data improves the model's reliability and minimizes the impact of outliers, leading to more accurate salary predictions.

4 Code

```
Module  vedic_mult_module (

 input [3:0] x,          // 4-bit multiplier x

 input [14:0] m,         // 15-bit operand m

output reg [31:0] result // 32-bit result);

 reg [31:0] partial_products [3:0]; // Array to store partial products


integer i;

 always @*

   begin

     for (i = 0; i < 4; i = i + 1)

begin

  partial_products[i] = m * (x & (1 << i)); // Calculate partial products

 end

 end

 // Calculate final result by adding all partial products

 always @*

  begin

 result = 0;
```

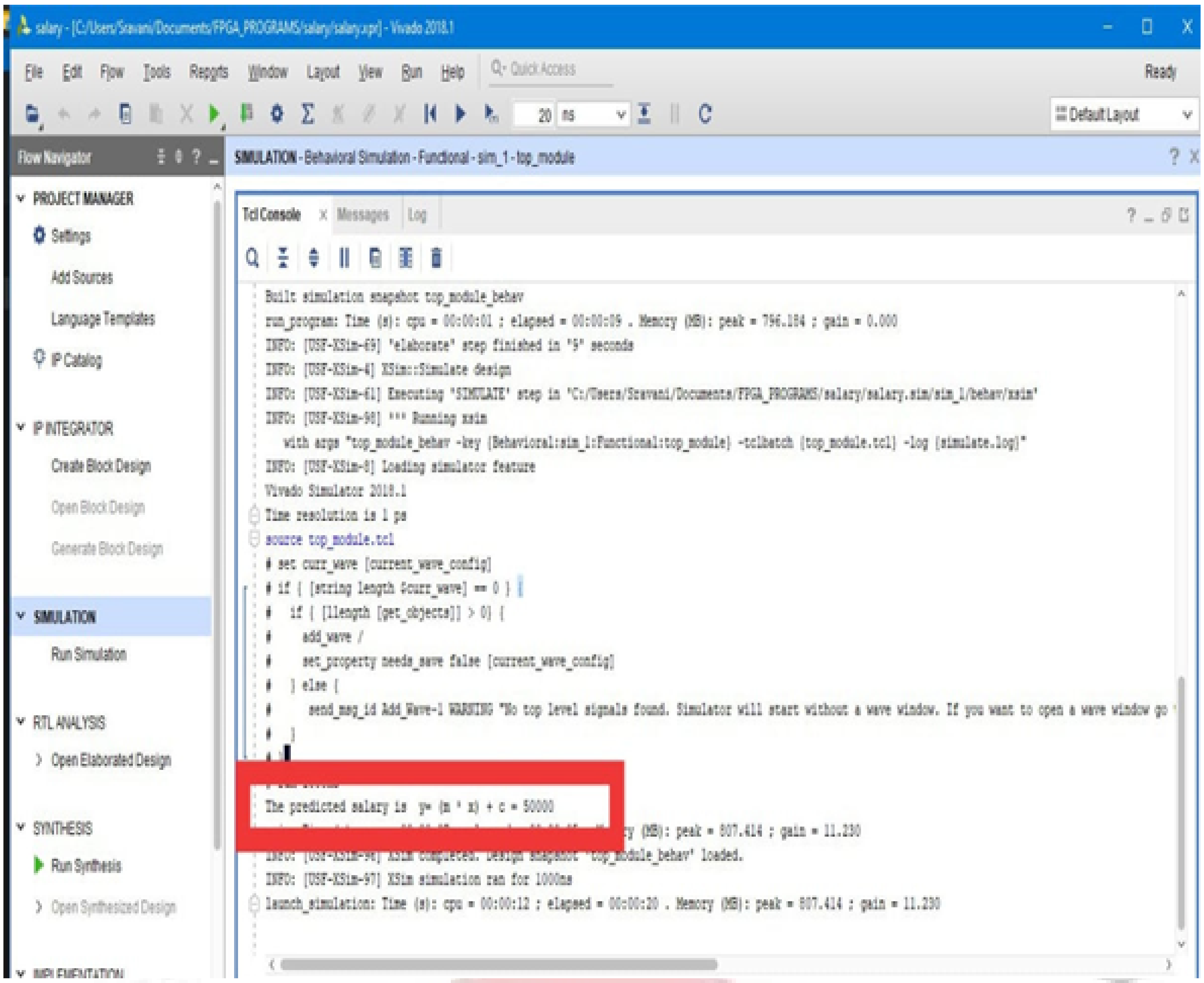
```
        for (i = 0; i < 4; i = i + 1)
            begin
                result = result + partial_products[i]; // Accumulate partial products
    end
Endmodule
module adder_module (
 input [31:0] operand1,   // 32-bit input operand
input [14:0] operand2,   // 15-bit operand c
output reg [31:0] sum     // 32-bit output sum);
always @*
  begin
     sum = operand1 + operand2; // Addition operation
  end
Endmodule
module top_module (
output reg [31:0] y      // 32-bit output y);

// Constants
 parameter [14:0] m = 15'd10000;  // 15-bit constant operand m
parameter [14:0] c = 15'd20000;  // 15-bit constant operand c
 parameter [3:0] x = 4'b0011;     // 4-bit constant multiplier x
 wire [31:0] mult_result;  // Wire to connect multiplication result
 wire [31:0] add_result;   // Wire to connect addition result
// Instantiate Vedic multiplication module
vedic_mult_module multiplier (
```

```
      .x(x),
```

```
      .m(m),
      .result(mult_result)   );
// Instantiate Addition module
 adder_module adder (
     .operand1(mult_result),  // Input operand1 is the multiplication result
     .operand2(c),           // Input operand2 is constant c
     .sum(add_result)   );     // Assigning output
 always @*
   begin
       y = add_result;  // Output is the final addition result
  end
   always @*
      begin
          $display("The predicted salary is  y= (m * x) + c = %0d", y);
     end
end module
```

5. Output



6. FUTURE SCOPE

In future iterations, this project can expand by integrating big data sources and advanced analytics techniques, such as natural language processing for extracting insights from job postings and economic indicators. Enhanced feature engineering could explore domain-specific metrics and time-series trends to better capture salary determinants. Implementing ensemble methods like Random Forest and model stacking could further improve prediction accuracy and robustness. Additionally, leveraging deep learning for feature representation and real-time predictions would enhance the model's capabilities in dynamic environments. Ethical AI considerations, including fairness metrics and explainable AI techniques, could ensure transparency and mitigate biases in predictions. Collaborating with industry experts would provide valuable insights for refining the model and validating its applicability in diverse organizational contexts.

7. Conclusion

Based on the enduring strengths of linear regression in salary prediction— its simplicity, interpretability, and historical significance— it is evident that while more complex models exist, linear regression remains pivotal. Future advancements in data collection, refining models, and ethical considerations will be crucial. Enhancements such as integrating advanced feature engineering techniques, exploring ensemble methods for improved accuracy, and incorporating ethical AI principles for fairness will further solidify its relevance. As the landscape evolves, the foundational role of linear regression in providing transparent and effective salary predictions will continue to shape the trajectory of predictive modeling in HR and economic domains, ensuring robust, interpretable insights for decision-makers.

8. References

1.Tzanis, George, et al. "Modern Applications ofMachine Learning." Proceedings of the 1st Annual SEERC Doctoral Student Conference–DSC. 2006.

2. Horvitz, Eric. "Machine learning, reasoning, and intelligence in daily life: Directions and challenges."Proceedings of. Vol. 360. 2006.

3. Mitchell, Tom Michael. The discipline of machine learning. Carnegie Mellon University, School ofComputer Science, Machine Learning Department,
2006.

4. Arum, R. (1998). The effects of resources on vocational student educational outcomes: Investeddollars or diverted dreams? Sociology of Education,
71, 130-151.

5. Lewis, C. D., 1982. Industrial and Business Forecasting Methods, London, Butterworths.