



# Predicting Annual Household Income Using Machine Learning

A Comprehensive Analysis

Photo by Pexels

# Table of Contents

01	Business Understanding
02	Problem Context
03	Project Objective
04	Project Goals
05	Data Overview
06	Methodology
07	Model Development
08	Results
09	Feature Importance
10	Conclusion

# Business Understanding

## Key Factors

- Understanding factors influencing household income is crucial for stakeholders.
- Policymakers, businesses, and researchers are primary stakeholders.
- Traditional methods may lead to inaccurate assessments and decisions.
- Machine learning offers new perspectives and more accurate predictions.

# Problem Context

## Dataset & Challenges

- Synthetic dataset simulates demographic and socioeconomic factors.
- Traditional analysis methods are limited and might not capture the full picture.
- Challenges include data quality, feature selection, and model accuracy.
- Accurate insights are vital for informed decision-making.



# Project Objective

## Main Aim

- Leverage machine learning to predict annual household income.
- Aim to gain more accurate insights for better decision-making.
- Inform policy decisions for socioeconomic improvements.
- Targeted marketing strategies based on income predictions.

# Project Goals

## Key Goals

- Develop a predictive model for annual household income.
- Validate the model's accuracy and reliability with metrics.
- Understand relationships between different features and income levels.
- Use insights to inform stakeholders and drive decisions.

# Data Overview

## Dataset Description

- Overview of the synthetic dataset used in the project.
- Features include various demographic and socioeconomic factors.
- Visualize the dataset with tables and graphs.
- Data quality and integrity are critical for model accuracy.



# Methodology



## Approach

- Data preprocessing includes handling missing values and encoding variables.
- Model selection involves evaluating different algorithms.
- Cross-validation used for reliable model evaluation.
- Feature scaling applied to normalize data for better model performance.



# Model Development

## XGBoost

- XGBoost selected for its superior performance.
- Data split into training and testing sets for model development.
- Evaluation metrics include accuracy, F1-score, and RMSE.
- Training process emphasizes model robustness and reliability.

# Results

## Performance

- Model accuracy and other metrics are presented.
- Graphs showing predicted versus actual income illustrate results.
- Key findings highlight model's predictive power.
- Insights derived from the model's predictions are discussed.



Photo by Pexels

# Feature Importance

## ■ Top Influencers

- List and describe the most influential features in the model.
- Visualization includes a bar chart showing feature importance.
- Insights into which factors most affect household income.
- Implications for stakeholders based on feature importance.



# Conclusion

## Summary

- Recap of the project objectives and key findings.
- Discuss implications of findings for different stakeholders.
- Highlight the importance of machine learning in socioeconomic research.
- Future work includes refining the model and exploring new data.