

Developing an Income Prediction Model using KLN and Bivariate Analysis

Murphy Efetobore Elo, Praise Olamide Abimbola

Student ID: x21220263@student.ncirl.ie, x22101012@student.ncirl.ie,

Cloud Machine Learning, MSc in Cloud Computing

National College of Ireland

Dublin, Ireland.

Project Problem

Many people have difficulty estimating their potential income based on their qualifications and skills, which can be a hindrance in career planning and decision-making. This project aims to develop a machine-learning model that can accurately predict income based on certain parameters such as age, education, skills, job title, industry, and so on. The model will help individuals estimate their potential income and guide career planning and decision-making.

Project Question:

Can we build a machine learning model that accurately predicts income for job seekers based on age, education, skills, and other parameters, and how well does it perform?

Proposed Solution:

We propose to develop a machine learning model for job seekers that predicts income based on various parameters such as age, education, skills, and other relevant parameters. The model will be trained on KLN (Kolmogorov-Smirnov Test) and bivariate analysis to identify the most relevant factors that affect income. The model will use these factors to predict income and provide recommendations to improve income potential.

Methodology

1. Data Collection (Murphy):

The dataset for this research will be collected from various sources such as government databases, surveys, and online job platforms. The dataset will include information about individuals such as age, education, skills, job title, industry, and income.

2. Data Cleaning and Preparation (Praise):

The collected data will be cleaned and prepared for machine learning by removing irrelevant information and handling missing values and outliers. We will also ensure that the data is properly formatted and encoded for use in the machine-learning model.

3. Feature Extraction (Praise)

We will use feature extraction techniques such as Principal Component Analysis (PCA) and feature scaling to extract the most relevant features that affect income. We will also perform correlation analysis to identify the relationship between different factors and income.

4. Machine Learning Model Development (Murphy):

We will develop a machine learning model based on techniques such as linear regression or decision trees to predict income based on the identified factors. We will use techniques such as cross-validation and hyperparameter tuning to ensure that the model is accurate and reliable.

Evaluation using Performance Metrics (Praise):

We will evaluate the performance of the machine learning model using performance metrics such as accuracy, precision, recall, and F1 score.

AWS Cloud Services:

We will be using Amazon Sagemaker as the cloud platform to create and train the model.

AWS S3 service will be used to store the dataset as an object store.

Overall, the objective of this research is to develop a machine-learning model that accurately predicts income based on various parameters such as age, education, skills, and other relevant factors. The proposed solution can be useful for individuals who want to estimate their potential income based on their qualifications and skills, and for employers who want to make informed decisions about salaries and compensation.