The Place to Start, Accelerate,
or Change your Career

IUNGO
solutions

STUDY CASE: Annual Income Prediction

A UK-based financial services company aims to develop a machine learning model to predict an individual's annual income. This model will be used to streamline loan application processes and offer personalised financial products. Click here for details about the dataset.

Project Outline:

1. Data Collection: The company has gathered data from 1,548 individuals, including demographic information, employment details, and financial status.
2. Data Pre-processing:
    - Handle missing values, particularly in the Type_Occupation column.
    - Convert categorical variables (e.g., GENDER, Car_Owner) into numerical format using techniques like one-hot encoding.
    - Normalise numerical features such as Birthday_count and Employed_days.
3. Feature Engineering:
    - Create new features like 'Age' from Birthday_count.
    - Develop a 'Financial_Responsibility_Score' based on Car_Owner, Propert_Owner, and Housing_type.
    - Create new features (e.g., Income per Family Member = Annual_income / Family_Members).
    - Convert Type_Income into binary categories (example: "Stable" vs. "Unstable" income).
4. Model Development:
    - Use Spark ML's Linear Regression or Random Forest Regression algorithms.
    - Split the data into training and testing sets.
    - Train the model using the input variables to predict Annual_income.
5. Model Evaluation:
    - Use metrics such as Mean Squared Error (MSE) and R-squared to assess model performance.
    - Employ cross-validation to ensure model robustness.
6. Hyperparameter Tuning:
    - Utilise Spark ML's ParamGridBuilder and CrossValidator for hyperparameter optimisation.