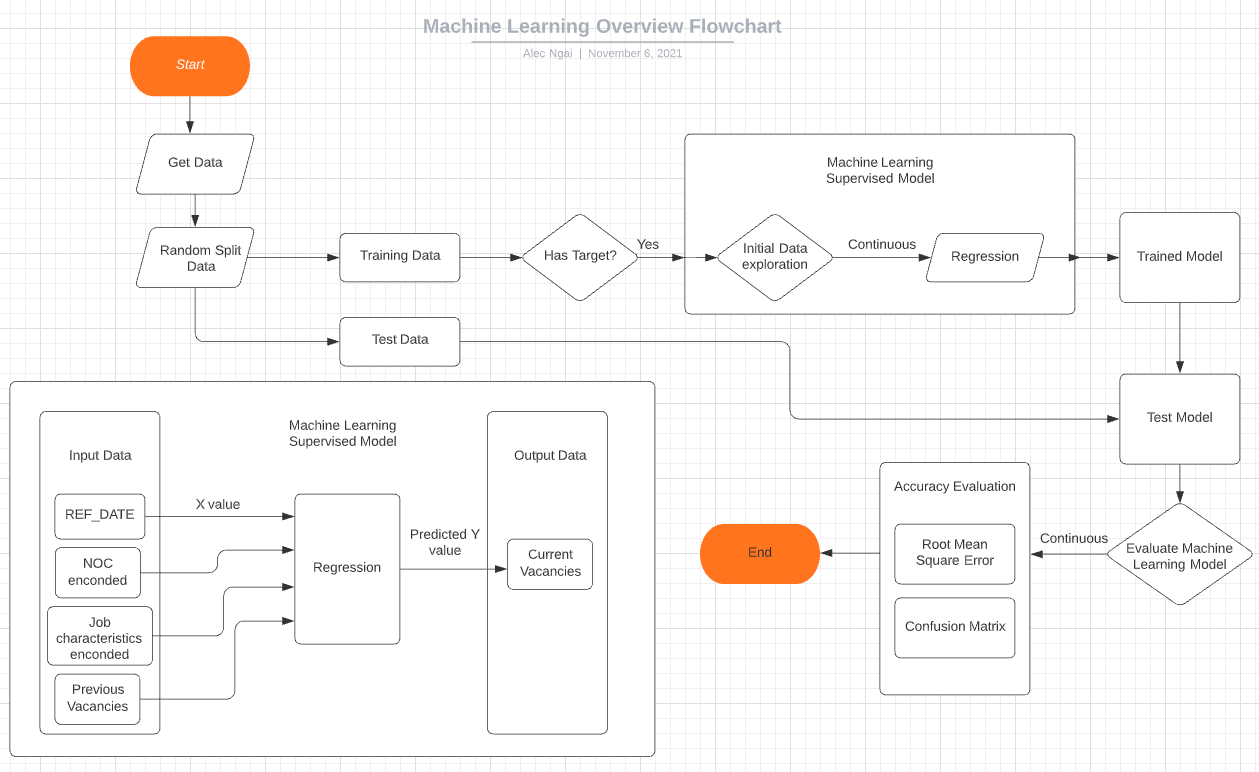
# **Machine Learning Model Mockup Description**



## **Description of data preprocessing:**

* Data Cleansing:
  + Raw data has a data quality index, the team decided to use data evaluated as Excellent, Very good, Good and Acceptable to have enough data to serve the project problem and questions.
  + Raw data has missing values for years 2020 and 2021. The team decided to drop these values as they may cause imbalance in the dataset.
  + Raw data contains three types of statistics, Job Vacancies, Proportion of Job Vacancies, Average Offered Hourly Wage, Machine learning Model will use Job Vacancies statistics as primary dataset. Other datasets will be used to enrich visualization and data interpretation.

* Data Transformation:
  + Several columns will need transformation, text extraction and encoding to be usable in the machine learning model
  + Unnecessary columns will be dropped.

## **Description of feature engineering and the feature selection including the team's decision-making process**

The team executed initial data exploration to review the current features available in the data set. Then evaluate what features can be most relevant to the project’s problem.

The team proposed to have the following features as input:

· Reference Date: quarterly periods when job vacancies existed

· National Occupancy Classification: The national classification of jobs (NOC)

· Job Characteristics: different Job characteristics and requirements (full time vs part time, education level required)

· Previous Job Vacancies: total number of vacancies available at a certain period.

The output feature will be:

Current Job vacancies: current total number of vacancies available

**Description of how data was split into training and testing sets**

The data will be split into training and testing randomly with stratification.

## **Explanation of model choice plan**

The project problem and desired output calls for a supervised machine learning model. The team will evaluate regression models using sample data and decide on the best model that fits the project’s purpose.