Machine Learning 1

Final Project

Regression and Classification

Prepared By:

1. Abdurehman Asfaw ---------- 448665
2. Bhupender Bhupender ---------- 466758

Datasets Description

`Delays` for **regression**

The dataset includes 1631327 observations in the training sample and 407832 in the test sample and the following columns:

- `Weekday` – Day of the week when the flight occurred (1 for Sunday, 7 for Saturday).
- `Month_of_Year` – The numerical month (1-12) when the flight took place.
- `Day_of_Month` – The day of the month (1-31) when the flight occurred.
- `Scheduled_Departure_Time` – The scheduled local time of flight departure.
- `Scheduled_Arrival_Time` – The scheduled local time of flight arrival.
- `Marketing_Airline` – The airline code under which the flight was marketed.
- `Marketing_Airline_DOT_ID` – Department of Transportation identifier for the marketing airline.
- `Flight_Number` – The number assigned to the flight by the operating airline.
- `Origin_Airport_ID` – Unique identifier for the departure airport.
- `Destination_Airport_ID` – Unique identifier for the arrival airport.
- `Flight_Cancelled` – Indicator of whether the flight was cancelled (1 = Yes, 0 = No).
- `Departure_State` – The state code of the departure location.
- `Arrival_State` – The state code of the arrival location.
- `Departure_Delay` – Total delay in minutes at departure.
- `Diverted_Airport_Landings` – Count of unplanned landings at other airports.
- `Taxi_Out_Time` – Time in minutes from gate departure until takeoff.
- `Taxi_In_Time` – Time in minutes from landing to gate arrival.
- `Flight_Diverted` – Indicator of whether the flight was diverted (1 = Yes, 0 = No).
- `Actual_Departure_Time` – The actual local time the flight departed.
- `Flight_Duration` – The duration of the flight in minutes from takeoff to landing.
- `Flight_Distance` – The total distance covered by the flight in miles.
- `Origin_Temperature` – The temperature at the origin airport at the time of the flight's departure.
- `Destination_Temperature` – The temperature at the destination airport at the time of the flight's arrival.
- `Origin_Wind_Speed` – The wind speed at the origin airport during the departure of the flight.
- `Destination_Wind_Speed` – The wind speed at the destination airport during the departure of the flight.
- `Origin_Precipitation` – The amount of precipitation, such as rain or snow, at the origin airport around the flight's departure time.
- `Destination_Precipitation` – The amount of precipitation, such as rain or snow, at the destination airport around the flight's arrival time.
- `Arrival_Delay` – Total delay in minutes at arrival (**outcome variable**, only in the training sample)

`job_change` for **classification**

Your task is to apply various ML algorithms (see the rules below) to build a model explaining whether a particular person **is willing to change job** based on the **training sample** and generate predictions for **all observations** from the **test sample**.

The dataset includes 12427 observations in the training sample and 3308 in the test sample and the following columns:

- `id` – unique observation identifier
- `gender` – gender of a person
- `age` – age of a person in years
- `education` – highest formal education level of a person attained so far
- `field_of_studies` – field of studies of a person
- `is_studying` – information whether a person is currently studying
- `county` – code of the county in which the person currently lives and works
- `relative_wage` – relative wage in the county of residence (as percentage of country average)
- `years_since_job_change` – years since a person last changed job
- `years_of_experience` – total number of years of professional experience of a person
- `hours_of_training` – total number of training hours completed by a person
- `is_certified` – does a person have any formal certificate of completed trainings
- `size_of_company` – size of a company in which a person currently works
- `type_of_company` – type of a company in which a person currently works
- `willing_to_change_job` – is a person willing to change job (**outcome variable**, only in the training sample)