**Mid Bootcamp Project Presentation 7 October 2022 Friday**

**Topic:** Analysis and prediction of average temperature based on meteorological data

**Slide 1:** Hello everyone. Today, I will discuss my mid bootcamp project about analysis and prediction of average temperature based on meteorological data.

**Slide 2:** Here is the outline of the presentation. Firstly, I will introduce the topic. Then, I will talk about the materials and methods. After that, I will show and discuss the results. Finally, I will conclude the presentation.

**Slide 3:** This study analyses and predicts the average temperature of two European cities, Dusseldorf and Munich in Germany using machine learning models. Dataset consists of meteorological data which ranges from 2000 to 2010.

The main objectives of this study include the following investigations:

* Analyzing relationships between meteorological parameters
* Hypothesis testing to decide whether the data sufficiently support the hypothesis
* Building linear regression and logistic regression models to predict the average temperature

In this case, the Dusseldorf data and Munich data are used to train and test the linear regression model, respectively. Additionally, logistic regression model is developed using Dusseldorf data.

**Slide 4:** Meteorological parameters in the dataset are average temperature, precipitation, humidity, cloud cover, wind speed, wind gust, global radiation, sunshine, pressure, and barbeque weather. Here, the precipitation and temperature data are observed from 2000 to 2010. The analyses and predictions are performed on .csv and excel files using Python, Mysql, and Tableau softwares.

The methods of this study include the followings:

• Get data

• Exploratory data analyses

• Data cleaning

• Data visualization

• Hypothesis testing

• Data wrangling

• Build model

One sided hypothesis is created to be tested. Based on average temperature in Dusseldorf between 2000-20, the hypothesis tests if average temperature in Dusseldorf between 2000-2010 is more than 10.8 C. P-value suggest that the null hypothesis can be rejected.

**Slide 5:** Here are the results of the linear regression model. Firstly, the model is trained using Dusseldorf data between 2000 and 2010. The predicted temperature for Dusseldorf is 11.10 °C. And the r2 score is 0.77. After that, the model is tested with Muenchen data for 2001. The predicted annual temperature for Muenchen is 11.50 °C. The r2 score of test is a little bit lower than that of the train. Because the model is tested with data that the model has not seen before. Our model predicts the average temperature well. The r2 score decreased up to 0.74 since the meteorological parameters making very high correlation were removed from the model such as minimum and maximum temperatures.

**Slide 6:** Logistic regression model is applied for Dusseldorf data. A new data was used for this purpose. This data gives information if weather is suitable to have outdoor activites such as barbeque. The data shows that people can’t make barbeque outside in 75% of the days between 2000-2010. This implies that the data is unevenly distributed. In this case, the data is manipulated to make it evenly distributed. The logistic regression model is applied before and after manipulation. The model score is 0.93 before manipulation. And high fraction of calculated performance metrics here suggests that the model accurately predicts if the weather is suitable outside for activities such as barbeque.

**Slide 7:** Here are the predicted temperatures in Muenchen for 2001 and their relationship with sunshine. The temperature increases with increasing sunshine during summer.

**Slide 8:** The average temperatures were analyzed and predicted for Dusseldorf and Muenchen cities based on meteorological parameters using linear and logistic regression models. The linear regression model predicts that average temperature for Dusseldorf between 2000-2010 is 11.10 °C. And Muenchen has annual temperature of 11.50 °C in 2001. Logistic regression model is useful before making plans outside as it classifies if the weather is suitable for outdoor activities or not.

**Slide 9:** Thank you very much for listening. If you have any comments questions, I am happy to answer.