Predicting Rains in Australia

Question/need:

- What is the framing question of your analysis, or the purpose of the model/system you plan to build?

Australia's weather has made headlines again this year. Parts of the country faced massive flooding because of heavy rainfalls. Many lost homes and lives as a result of this unexpected calamity. Not long ago, the country faced devastating forest fires which burnt tens of millions of acres. Because of these contrasting weather conditions, Australian Government Bureau of Meteorology has approached me to build a better model that can predict rainfall.

For the first phase, I will be working on 10 year Australian climate data to come up with a model that can predict whether there will be rainfall for the next day over different parts of the country.

Future phases will include predicting the amount of rainfall.

- Who benefits from exploring this question or building this model/system?

Weather impacts all sectors of an economy either directly or indirectly. Predicting precise weather conditions might help prevent life and property damage. Moreover, any key findings might be applicable to other countries which are drought prone like India.

Data Description:

- What dataset(s) do you plan to use, and how will you obtain the data?

Australian weather dataset which recorded 10 years of weather observations. Data is downloaded from Kaggle:
https://www.kaggle.com/jsphyg/weather-dataset-rattle-package

Dataset contains about 140,000 rows and 22 features. Data can be downloaded as a CSV file and read into pandas using the pd.read_csv function.

- What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?

Each row records weather conditions such as date, location, min/max temperatures, wind speed, evaporation, sunshine and other weather characteristics.

Target variable is whether it will rain tomorrow or not.

- If modeling, what will you predict as your target?

Classification modeling techniques like logistic regression, KNN, Decision Trees and XG Boost

Tools:

- How do you intend to meet the tools requirement of the project?
    - SK Learn will be used for modeling
- Are you planning in advance to need or use additional tools beyond those required?
    - Tableau for visualization

MVP Goal:

- What would a minimum viable product (MVP) look like for this project?

MVP will have a cleaned dataset with EDA and a baseline classification model without tuning hyperparameters.