TASK A

Deepak Kumar Cheranjeev Rama
ppxdc3@nottingham.ac.uk

To address the challenge faced by a French bakery in optimizing production and reducing waste, a project was initiated to develop a predictive dashboard using historical weather and sales data. The aim was to create a tool that predicts average temperature (`tavg`) and suggests appropriate production quantities based on past sales patterns.

Two datasets were utilized: weather data from January 1, 2021, to October 31, 2022, and bakery sales data from 2021 to 2022. The weather dataset includes various meteorological variables, while the sales dataset contains transaction records of different bakery items. Data cleaning involved converting the `date` columns in both datasets to `datetime` format for consistency and extracting `day`, `month`, and `year` from the `date` column in the weather dataset to facilitate merging and analysis.

The decision was made to use only the average temperature (`tavg`) for several reasons. First, `tavg` is a comprehensive measure that captures the overall climate condition of a day, which is likely to influence consumer behavior more consistently than other weather metrics. Second, using `tavg` simplifies the model, making it easier to interpret and less prone to overfitting. Finally, other temperature-related features like `tmin` and `tmax` are highly correlated with `tavg`, and including all could introduce multicollinearity, complicating model training.

A Random Forest Regressor was chosen for its robustness and ability to handle non-linear relationships between the input features (day, month, year) and the target variable (`tavg`). The dataset was split into training and testing sets using an 80-20 split to ensure the model's performance could be evaluated on unseen data. Bayesian Optimization was used to find the best hyperparameters for the Random Forest model, ensuring optimal performance. The model was evaluated using Mean Absolute Error (MAE) to quantify the prediction accuracy. The model achieved 1.55 MAE, making it highly accurate, shown in figure 1.

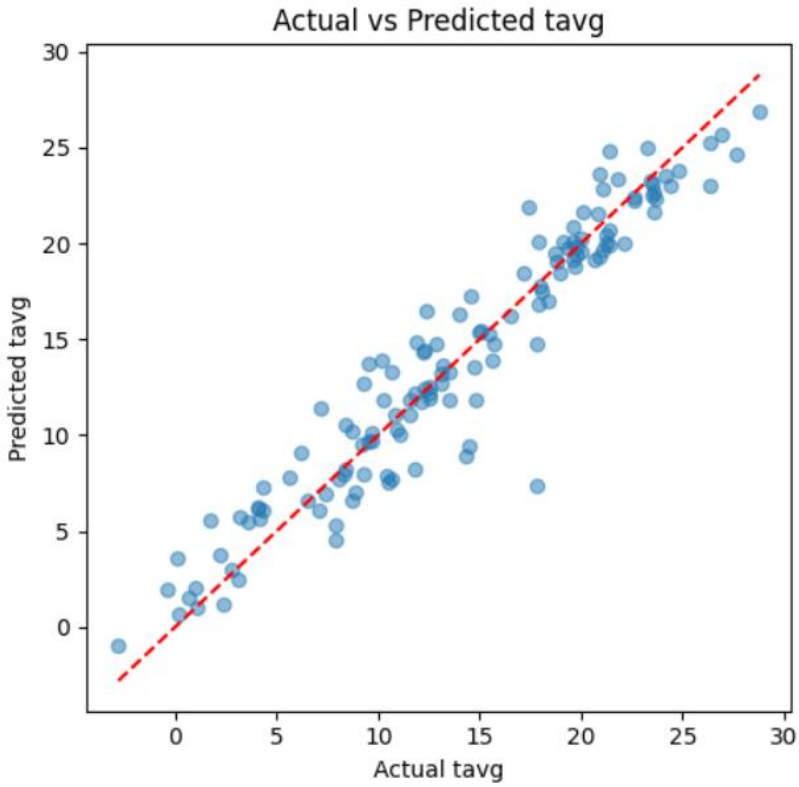

Figure 1

The dashboard allows users to input a specific date (day, month, year), and the model predicts the average temperature (`tavg`) for that date. The predicted temperature is then used to find historical sales data with similar temperatures to suggest production quantities. Streamlit was used to create an interactive web interface where users can input the date. The trained Random Forest model predicts the `tavg` for the input date. Historical data is filtered to include only records from the same month as the input date. The historical date with the temperature closest to the predicted `tavg` is identified, and the sales data from that date is used to suggest production quantities.

The dashboard displays the suggested production quantities using an interactive bar chart, implemented with Plotly. This visualization helps bakery operators quickly understand the recommended quantities for different bakery items, facilitating informed decision-making. The final Random Forest model was saved using `joblib` for efficient loading and prediction within the Streamlit app, ensuring that the dashboard can make real-time predictions based on user input, as shown in figure 2.

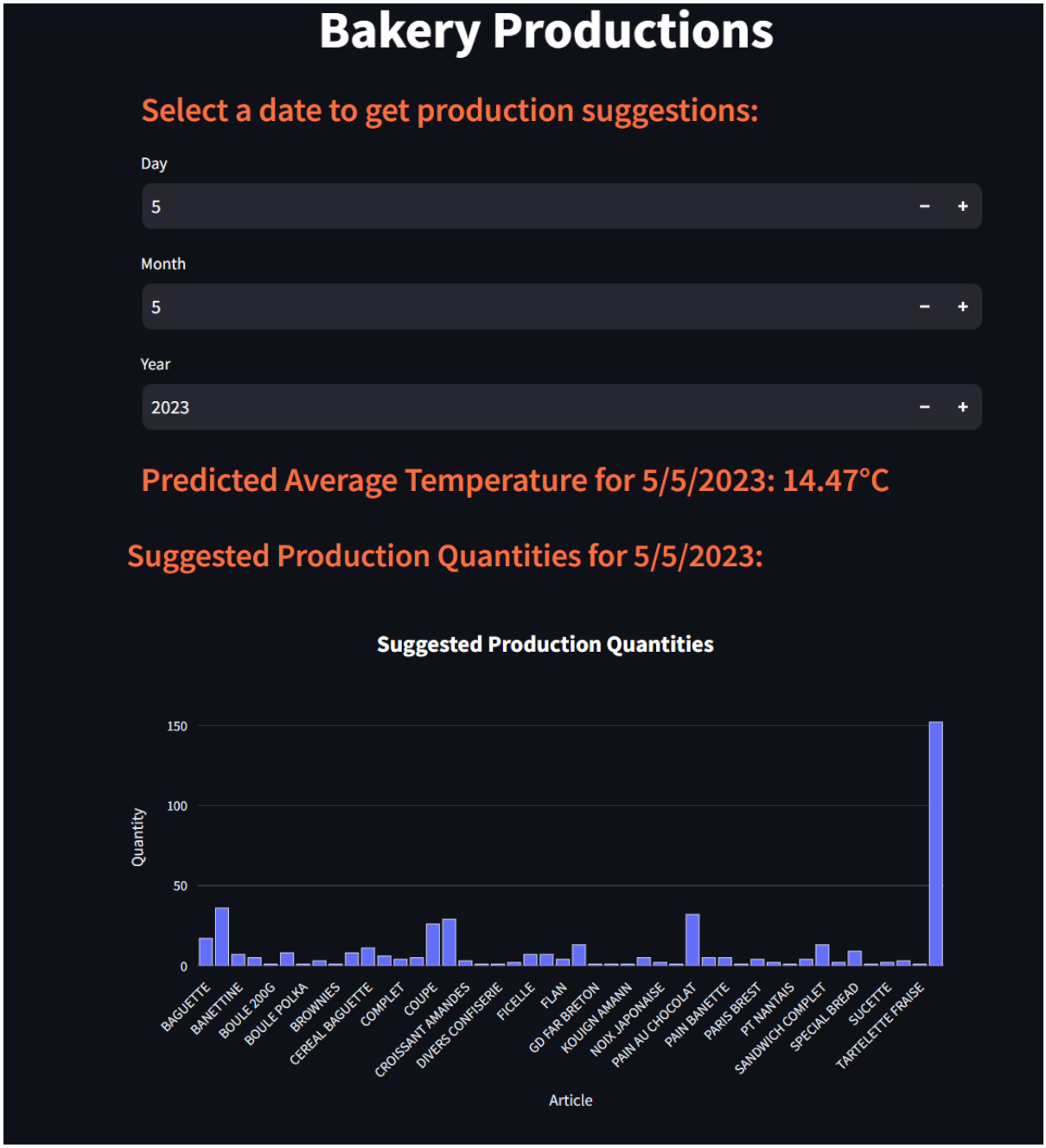

Figure 2

This project successfully integrates machine learning with an interactive web application to provide valuable insights for bakery production planning. By focusing on average temperature (`tavg`), the model leverages a key weather variable to predict sales patterns, ensuring simplicity and interpretability. The dashboard serves as a practical tool for bakery operators, helping them optimize production based on weather forecasts and historical sales data. The next steps could involve incorporating additional features, such as precipitation or special events, to further refine the model's predictive power and expand its applicability.