# Analysis of Flood Risk, Median Prices, and Historical Flood Data

**1. Regression/Classification Methods Used**

- **XGBoostRegressor/Classifier**: Boosting techniques were used to iteratively refine errors. Hyperparameter tuning (e.g., learning_rate=0.27, max_depth=14) optimized performance.
- **Random Forest with SMOTE**: Used exclusively in Model 3, where it outperformed XGBoost by 2% in recall. However, it was rejected for other models due to longer training times.
- **Other Methods Investigated but Rejected**: KNN and Neural Networks were excluded due to poor compatibility with label-encoded data, reliance on numeric transformations, and distance-based calculations.

**2. Feature Importance**

**Key Features**:

- **Elevation**: Strongly correlated with both house prices and flood risk.
- **Distance to Watercourse**: Highly predictive for flood risk and historical flooding.
- **NearestWatercourse**: Contributed significantly to house price predictions.
- **Local Authority**: A critical determinant for flood risk classification.

**Unimportant Features**:

- **Households and Headcounts**: Variability due to policies like the Affordable Homes Programme reduced their utility for modelling.
- **Geographical Splitting**: Longitude-based splits for coastal influence were considered but not effective.
- **Interaction Terms**: Discarded as they did not significantly improve performance.

**Inclusion of Historical Flooding**: Adding historical flooding data, particularly with SMOTE, could improve flood-related predictions by addressing class imbalances.

**3. Data Visualization and Analysis**

**Risk Map and Web Tool**:An interactive platform visualizes flood risks and real-time rainfall, enabling users to identify vulnerable areas and assess immediate flooding risks.

**Key Insights**:

- Major flooding events correlate with high-intensity rainfall.
- On wet days, heavy rainfall covers the UK, with the south facing the highest risk of economic loss, emphasizing the need for close rainfall monitoring in this region.

**Conclusion**

This report highlighted the effectiveness of XGBoost and Random Forest models. Feature analysis revealed elevation and proximity to watercourses as critical predictors. Future improvements could integrate additional data, such as historical flooding, to further enhance model accuracy.

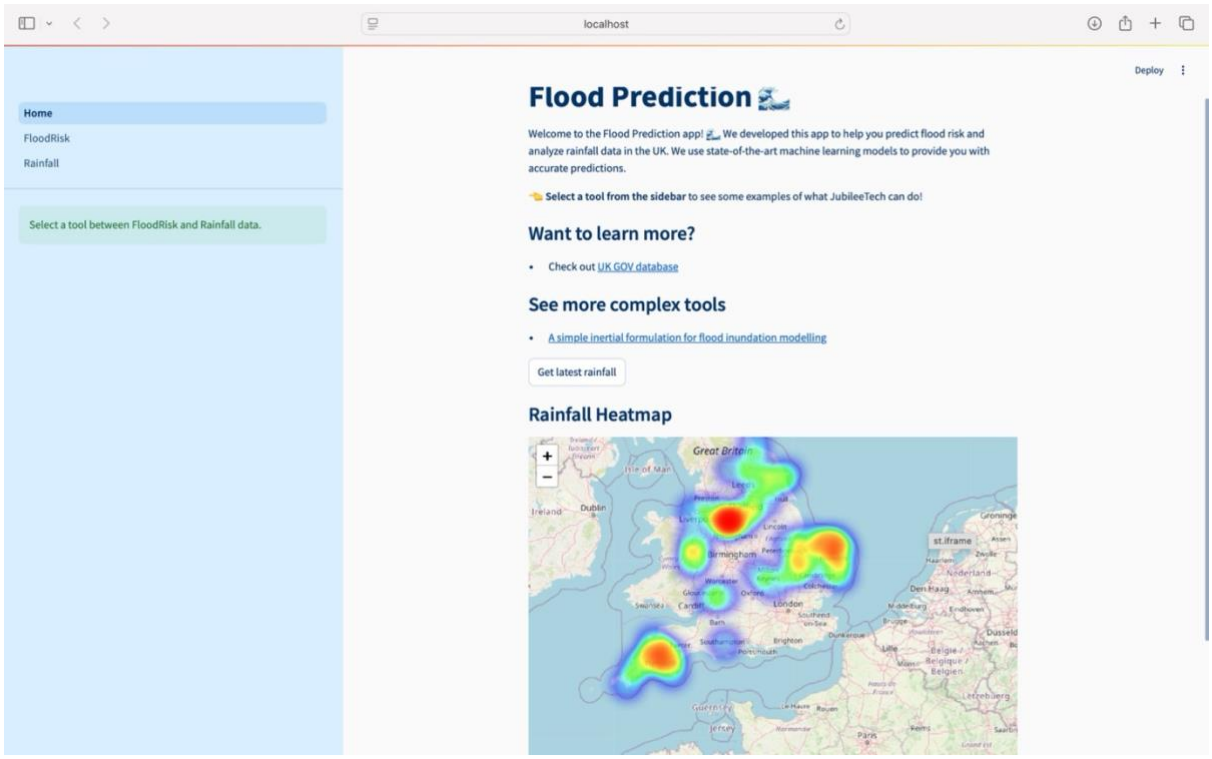


*Figure 1.Streamlit Home Page*

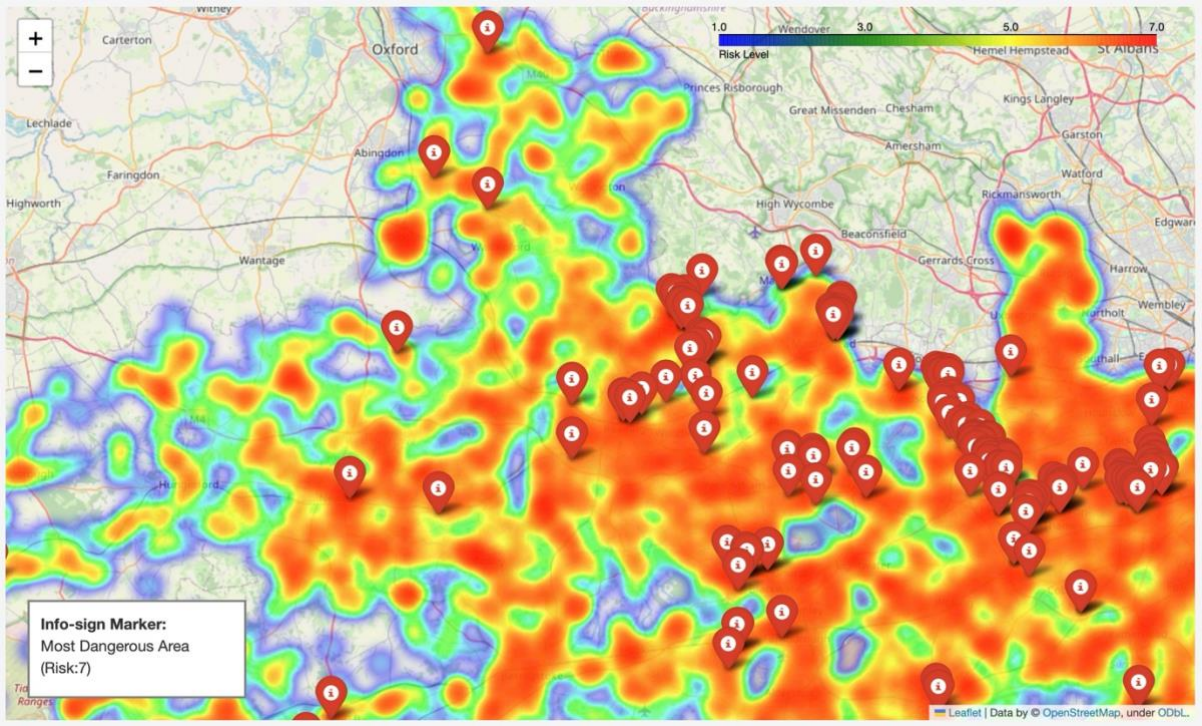


*Figure 2: Risk Map of Predicted Data*

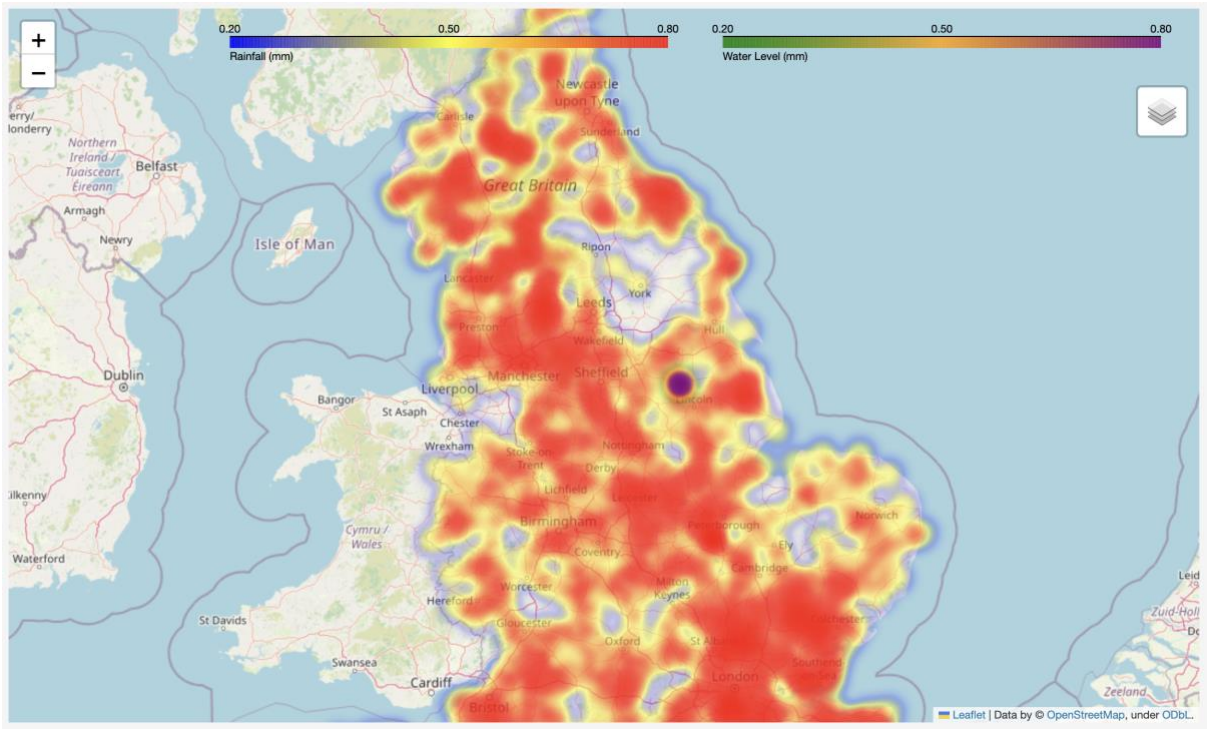


*Figure 3: Rainfall and Level Maps for Wet Day*

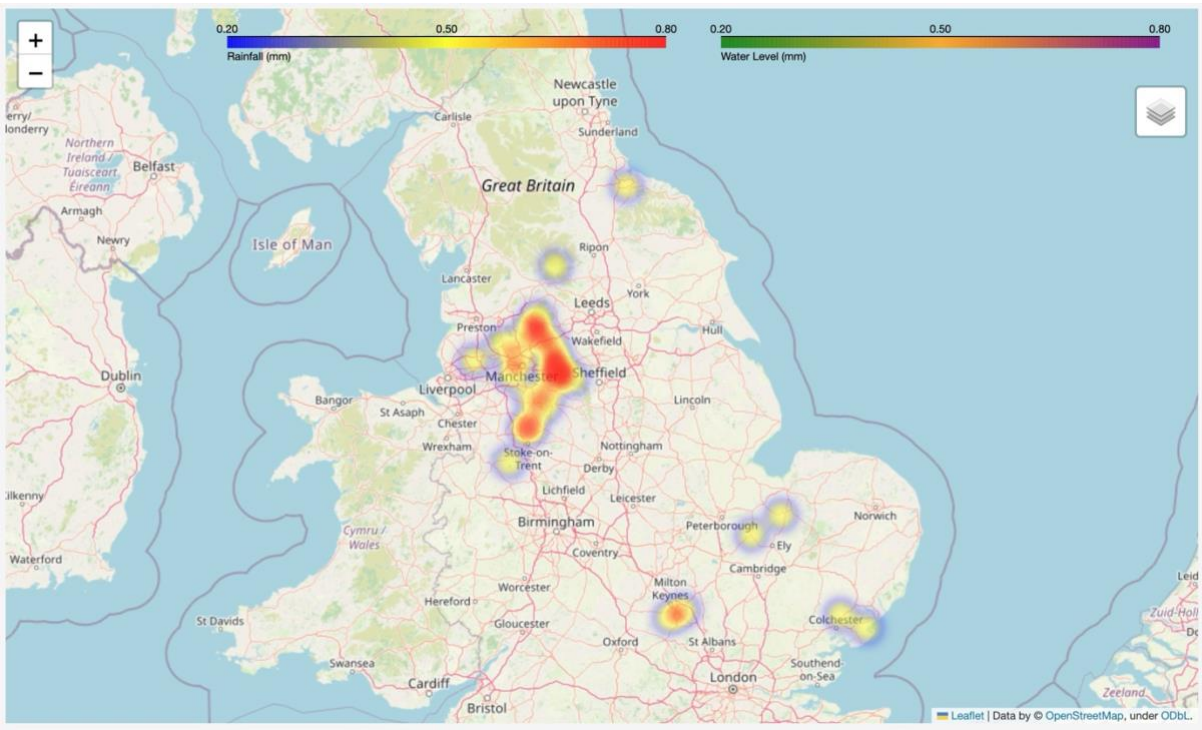


*Figure 4: Rainfall and Level Maps for Typical Day*