The Forestry and Agricultural sectors are key to the economy in South Africa, and to development in Africa. The biggest threats to the potential growth of these industries is climate change and introduced pests and diseases. The Forestry and Agricultural Biotechnology Institute and its research partner, the Institute of Commercial Forestry Research (ICFR), are at the forefront of helping the SA forestry and agriculture industries and related government departments to develop leading, appropriate technologies to address such issues. These institutes have an increasing need to more optimally integrate data from across these fields such as planting stock, pests and diseases with weather, soil and other data. In this project, we wish to unlock the potential of **70 years of weather data** to serve as a basis for such data integration. The project includes data appraisal and standardization, development of visualization and user interface application and analysis/visualization of patterns of variation that would make this invaluable dataset useable for both research and management purposes in future.

Enabling weather based decision making for forestry research and management.

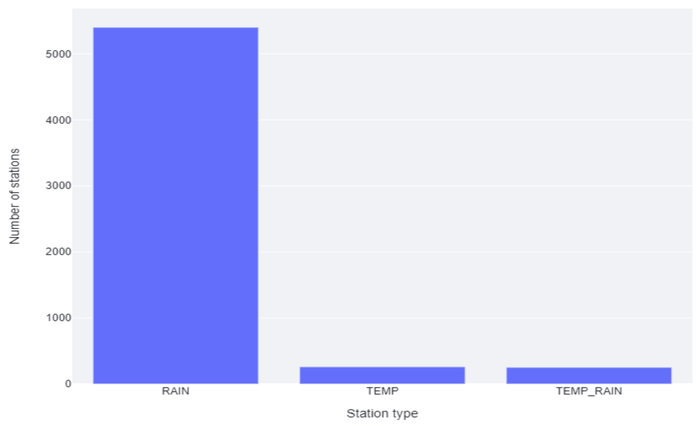
INTRO

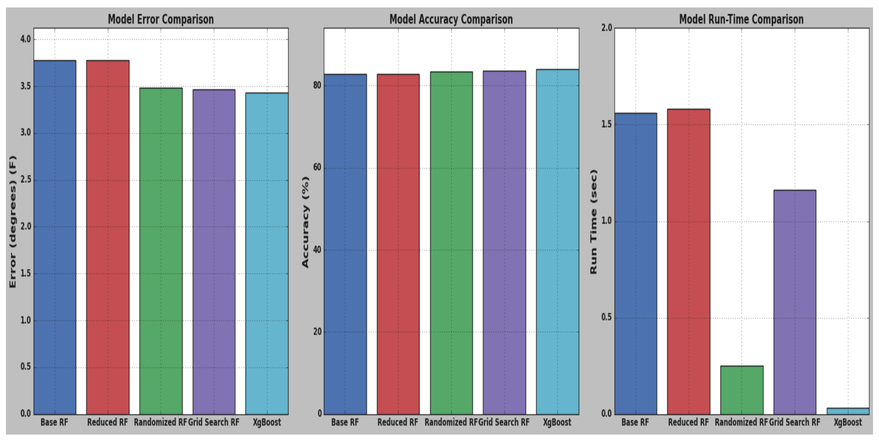
- The Institute of Commercial Forestry Research (ICFR) is the custodian of weather data on behalf of the forestry industry. The database consists of daily minimum and maximum temperatures and rainfall from 1950 to current. Part of the database (1950-2000) was stored in a user-friendly application, which is now obsolete.
- The remaining datasets are stores as text files. Currently the data cannot be readily accessed, visualized by potential users, and has not been analyzed for trends in variation over time.
- The aim of this project is to develop a web-based application for the storage, interrogation, visualization and retrieval of weather data. The application should also be able to perform basic analysis and have a certain degree of scenario modelling.
- The application functionalities must be user-friendly, accessible through the web for easy access by the industry, able to integrate into cloud-based data platforms, and updatable.

METHODS

1. Daily weather data collected from 6352 weather stations from the year 1950 to 2019 (107 million observations).
2. Data visualization performed based on the selection of weather station.
3. Prediction of **Rainfall, Minimum and Maximum temperature** was performed using random forest(RF) and extreme gradient boosting(XgBoost) algorithms.
4. Hyper-parameter tuning was performed on each of the algorithm to select the best parameter for the model.
5. The visualisation and the best random forest model and extreme gradient boosting was then deployed to the WebApp using streamlit package in Python.

RESULTS





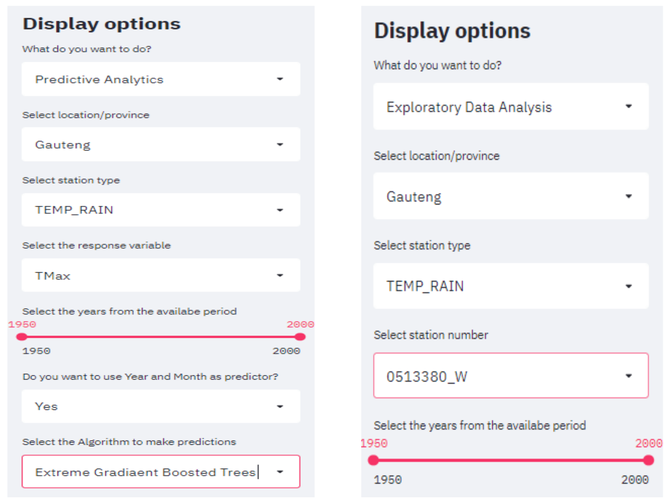
The model performance based on Gauteng province

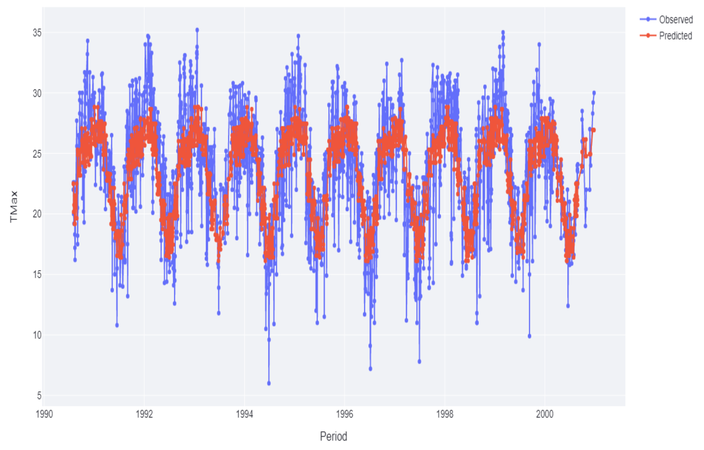
Model	MAE	MSE	RMSE
Random Forest	**2,95**	**14,01**	**3,74**
XgBoost	**2,69**	**11,75**	**3,43**

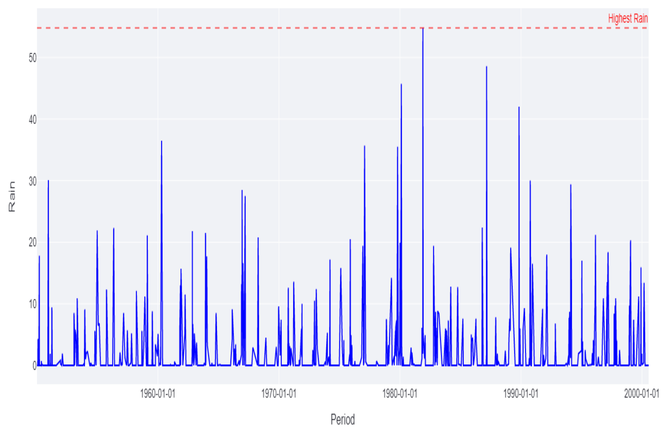
DISCUSSION

- The random forest model with random parameter search was found to have comparable accuracy as the other RF models but has a less run time.
- The extreme gradient boosting model on the other hand has similar accuracy measures but with less runtime than the random parameter search RF

Streamlit App Visualization







Ndivhuwo Mphephu

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Department of Computer Science

Faculty of Engineering, Built Environment and Information Technology

Fakulteit Ingenieurswese, Bou-omgewing en Inligtingtegnologie / Lefapha la Boetšenere, Tikologo ya Kago le Theknolotši ya Tshedimošo

Capstone Project - MIT 808

Course Coordinators:
Dr. Vukosi Marivate (vukosi.marivate@cs.up.ac.za)
Abiodun Modupe (abiodun.modupe@cs.up.ac.za)

Scan me