Forecast Tx

Senior Design

Department of Computer Science and Engineering

Sanchez, N., Berry, M., Phommalay, K., Patel, K., Simbakwira, K.

Background

Sponsored by State Farm, this project leverages machine learning to predict severe weather events. Specifically, thunderstorms and hail across Texas, with forecasts extending five years into the future. The predictive models are trained using over 70 years of historical data from ERA5 reanalysis and NOAA storm reports, capturing long-term weather trends and storm behavior.

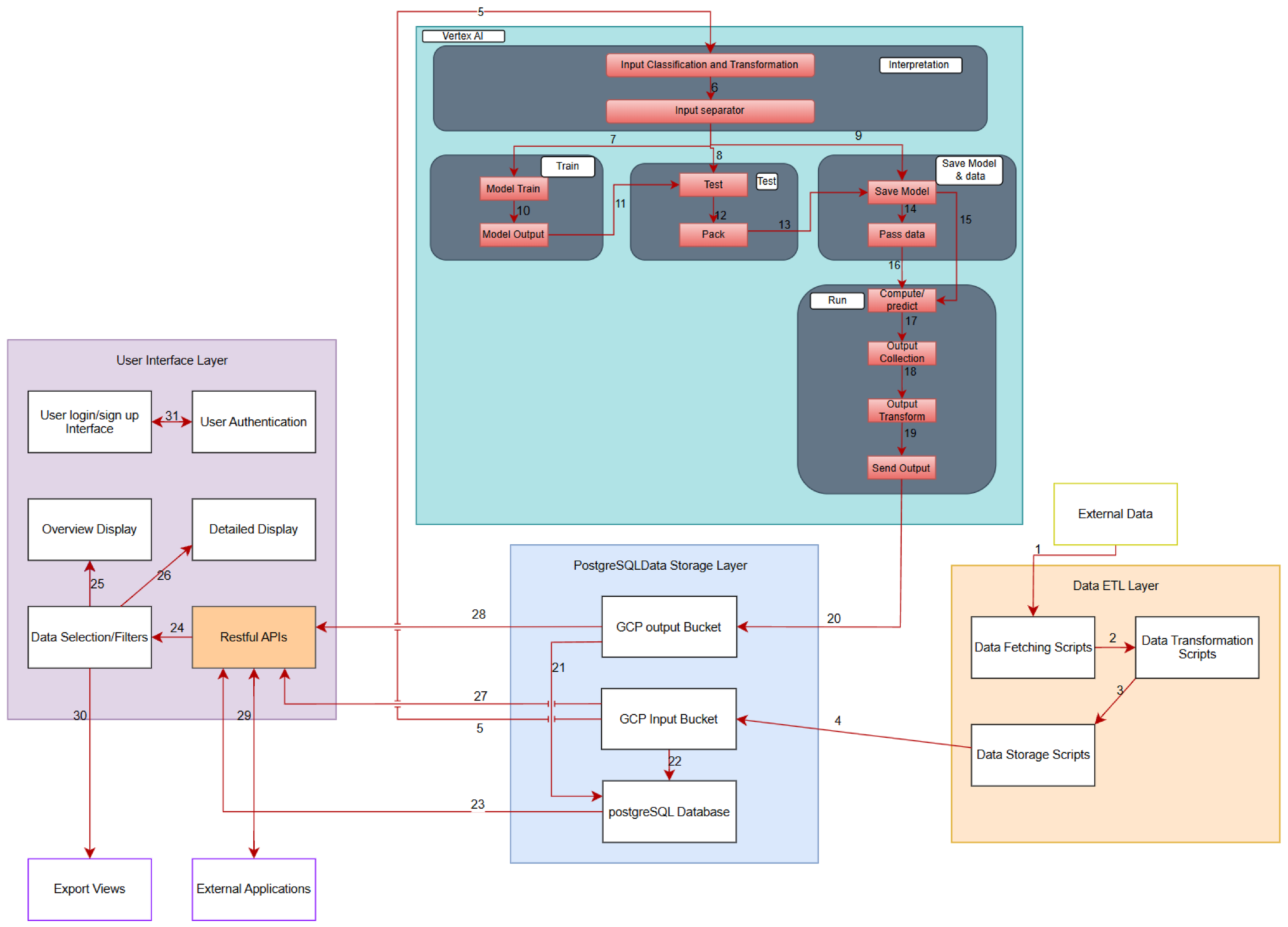
The results are presented in a user-friendly, Firebase-hosted web application featuring interactive maps and statistical visualizations. Users can explore predicted storm frequencies, locations, and intensities across different timeframes. The backend is supported by Google Cloud Platform services, including Cloud Storage, PostgreSQL databases, and virtual machines to enable scalable data processing and real-time API delivery.

Designed to support State Farm's operational planning, this system aims to assist in disaster preparedness, risk assessment, and other insurance-related decision-making tasks by providing a data-driven outlook on future severe weather activity in Texas.

Key Requirements

Severe weather events, especially hail, wind, and flooding, have become increasingly frequent and damaging across Texas. Such events are driving up insurance rates and threatening property insurability. In response to this growing concern, State Farm seeks advanced tools to better anticipate future events and minimize losses. Our project addresses this need by leveraging over 70 years of historical weather data from NOAA and other sources to train machine learning models that forecast the frequency and severity of weather-related perils in specific high-risk counties. These predictive insights can support proactive disaster preparedness, risk-based insurance modeling, and ultimately help protect lives and property. According to NOAA's National Weather Service, extreme weather events have trended upward in both intensity and cost, making predictive solutions not just useful, but essential.

Architectural Design



Architectural Layers

Layer	Description	Inputs	Outputs
Data ETL	This layer uses python scripts to handle large data sets and transform them for storage and model input.	ERA5 hourly weather attribute data in GRIB file format	Cleaned, organized, sorted CSV files for storage and model input
Data Storage	This layer consists of various forms of cloud storage used for holding model input and output data as well as data views in PostgreSQL to serve formatted data to the front end.	Transformed CSV files from the ETL layer and model outputs from the ML layer (JSON/CSV)	Relational tables and views in PostgreSQL and GCS files for fast query and visualization
Machine Learning	Provides a web-based platform with authentication, filters, and interactive views to explore model predictions and historical data.	Pre-processed data from storage layer	Trained models; prediction results in JSON/CSV format; model metadata
User Interface	This layer uses Vertex AI to train, validate, and run machine learning models using weather data. It also manages model output and stores results in the database.	Filtered query results, JSON/CSV files from PostgreSQL and GCS buckets via RESTful APIs	Interactive displays including maps, charts, and dashboards with exportable results and user feedback

Conclusions and Future Work

- Our project successfully predicts useful weather data and showcases how machine learning can be leveraged to provide valuable data to State Farm.
- The major areas that could be improved with additional work would be creating additional interactive user interfaces and improving machine learning model accuracy.
- By establishing a model improvement process, the ML model can be improved with re-training and testing, adjusting input feature weights and biases to produce a more accurate prediction. Establishing a clear pipeline to apply this iterative process would allow for time and resource effective model re-training to improve data accuracy and veracity as well as explore predictions of additional valued weather indicators.

Machine Learning Models

Model 1

Model 1 uses ConvLSTM (Long short-term memory) and 70 years of hourly weather attribute data from across Texas to predict weather attributes from 2025 to 2030.

Statistics

Mean Absolute Error (MAE): 0.1232 (~12%)

Model 2

Model 2 uses Classification and Linear Regression models. Weather attributes, as well as recorded hail and thunderstorm events from NOAA severe weather database, are used to predict weather events in Texas from 2025 to 2030.

Statistics

Overall **Accuracy:** 96.87%
Recall (90.68%)

Training Parameters

ERA5 Reanalysis Hourly Data Select Parameters Documentation

Short Name	Full Name	Units	Description
2t	Two-meter temperature	K	Air temperature at two metres above ground
2d	Two-meter dewpoint temperature	K	Dew point temperature at two metres above ground
sp	Surface pressure	Pa	Atmospheric pressure at the Earth's surface
tcc	Total cloud cover	0 – 1	Fractional coverage of all clouds in a grid box (0 = clear, 1 = overcast)
lcc	Low cloud cover	0 – 1	Fractional coverage of low-level clouds in a grid box
mcc	Medium cloud cover	0 – 1	Fractional coverage of mid-level clouds in a grid box
hcc	High cloud cover	0 – 1	Fractional coverage of high-level clouds in a grid box
10u	10 m U-wind component	$m\ s^{-1}$	Eastward wind component at 10 m above ground
10v	10 m V-wind component	$m\ s^{-1}$	Northward wind component at 10 m above ground
100u	100 m U-wind component	$m\ s^{-1}$	Eastward wind component at 100 m above ground
100v	100 m V-wind component	$m\ s^{-1}$	Northward wind component at 100 m above ground
cape	Convective available potential energy	$J\ kg^{-1}$	Potential energy available to an ascending air parcel for convection
vit	Vertical integral of temperature	$K\ kg\ m^{-2}$	Vertically integrated atmospheric temperature over the full column
vike	Vertical integral of kinetic energy	$J\ m^{-2}$	Vertically integrated atmospheric kinetic energy over the full column
vitoe	Vertical integral of total energy	$J\ m^{-2}$	Vertically integrated total energy (internal + potential + latent + kinetic) over the full column
tclw	Total column cloud liquid water	$kg\ m^{-2}$	Liquid water path: total liquid water in cloud droplets integrated over the atmospheric column
tciw	Total column cloud ice water	$kg\ m^{-2}$	Ice water path: total ice water in cloud ice crystals (and snow) integrated over the atmospheric column

References

1. Google, Inc., "Google Cloud Platform Documentation," *Google Cloud Platform*, 2024. [Online]. Available: https://cloud.google.com/docs
2. Overleaf, "Collaborative Writing and Publishing Tool for LaTeX," *Overleaf*, 2024. [Online]. Available: https://www.overleaf.com
3. Agafonkin, V., "Leaflet: An Open-Source JavaScript Library for Interactive Maps," *Leaflet.js*, 2024. [Online]. Available: https://leafletjs.com
4. Hersbach, H., et al., "ERA5 hourly data on single levels from 1940 to present," *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*, 2023. [Online]. Available: https://cds.climate.copernicus.eu
5. NOAA National Centers for Environmental Information, "Storm Events Database," *National Oceanic and Atmospheric Administration (NOAA)*, 2024. [Online]. Available: https://www.ncdc.noaa.gov/stormevents/textsearchhelp.jsp
6. Google, Inc., "Firebase Hosting Documentation," *Firebase by Google*, 2024. [Online]. Available: https://firebase.google.com/docs/hosting

Summer 2025