COURSE PROJECT

Hurricane IDALIA Path Prediction

DMS672: DATA MINING & KNOWLEDGE DISCOVERY

SUMMER 2025

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

Submitted by

Abhiraj Akhouri (218170032**)** **Subhadip Baidya (**221092**)**

Aditya Dinesh Durgapal (240055**)** **Rohak Debnath (**220905**)**

Prof. Faiz Hamid

Department of Management Studies

Indian Institute of Technology Kanpur

Hurricane IDALIA Path Prediction Project Report

1. Hurricane IDALIA

Hurricane Idalia was a significant Atlantic hurricane event in August 203, tracked and analyzed using detailed GIS data from the National Hurricane Center and related sources. The project focuses on predicting the path of Hurricane Idalia using advanced data mining and machine learning techniques, leveraging both historical and real-time geospatial data.

2. Data

Source: NHC database

2.1 GIS Nature

- Spatial Data: All layers are GIS-enabled, containing geometric columns (POINT, LINESTRING, POLYGON) for mapping and spatial analysis.
- Temporal Data: Each record includes time-related fields (e.g., ADVDATE, TAU) to enable time series modeling.
- Meteorological Attributes: The _5day_pts layer includes wind speed, pressure, and other storm characteristics.

2.2 Layers and Columns

The hurricane dataset is organized into 4 GIS layers, each representing different aspects of the storm's forecast and impact:

Layer	Geometry Type	Columns	Description
_5day_lin	LINESTRING	STORMNAME, STORMTYPE, ADVDATE, ADVISNUM, STORMNUM, FCSTPRD, BASIN, geometry	Forecast path lines
_5day_pgn	POLYGON	STORM NAME, STORMTYPE, ADVDATE, ADVISNUM, STORMNUM, FCSTPRD, BASIN, geometry	Forecast cone polygons (uncertainty regions)
_5day_pts	POINT	ADVDATE, ADVISNUM, BASIN, DATELBL, DVLBL, FCSTPRD, FLDATELBL, GUST, LAT, LON, MAXWIND, MSLP, TAU, TCDIR, TCSPD, geometry	Forecast points with meteorological attributes

_ww_wwlin	LINESTRING	STORMNAME, STORMTYPE, ADVDATE, ADVISNUM, STORMNUM, FCSTPRD, BASIN, TCWW, geometry	Watch/warning lines

Layer: _5day_lin - Number of rows: **56**
Layer: _5day_pgn - Number of rows: **56**
Layer: _5day_pts - Number of rows: **504**
Layer: _ww_wwlin - Number of rows: **243**

2.3 Variable Description

Variable	**Description**	**Data Type**	**Example Value**
STORMNAME	Name of the storm (if named)	String	"IDALIA"
STORMTYPE	Type of storm (eg TD: Tropical Depression.)	String	"HU"
ADVDATE	Advisory date and time (local time, string format)	String	"400 PM CDT Sat Aug 26 2023"
ADVISNUM	Advisory number or identifier ()	String/Int	"1", "1A"
STORMNUM	Storm number	Int/Float	10.0
FCSTPRD	Forecast period (hours into the future)	Float	120.0
BASIN	Ocean basin code (e.g., AL for Atlantic)	String	"AL"
geometry	Geometric shape (POINT, LINESTRING, POLYGON) representing the spatial aspect	Geometry Object	POINT(-86.1 21.1)
TCWW	Type of tropical cyclone watch/warning	String	"TWA"

_5day_pts Layer Specific

Variable	Description	Data Type	Example Value
DATELBL	Date label (human-readable, often matches ADVDATE)	String	"4:00 PM Sat"
DVLBL	Development label (storm intensity, e.g., "D" for Depression)	String	"D"
FLDATELBL	Forecast label (date and time for the forecast point)	String	"2023-08-26 1:00 PM Sat CDT"

GUST	Maximum wind gust at this forecast point (knots)	Float	35.0
LAT	Latitude of the forecast point (degrees N, negative for S)	Float	21.1
LON	Longitude of the forecast point (degrees E, negative for W)	Float	-86.1
MAXWIND	Maximum sustained wind speed at this point (knots)	Float	40.0
MSLP	Minimum sea level pressure at this point (hectopascals, hPa)	Float	1005.0
TAU	Forecast hour (lead time from advisory, in hours)	Float	0.0, 12.0, 24.0
TCDIR	Storm direction (degrees from North; 9999 if missing)	Float	360.0, 9999.0
TCSPD	Storm movement speed (knots; may be 9999 if missing)	Float	0.0, 9999.0
TIMEZONE	Time zone of the advisory	String	"CDT"
VALIDTIME	Valid time for the forecast point (string, e.g., "26/1800")	String	"26/1800"
SSNUM	Saffir-Simpson storm number (category)	Float	1.0
STORMSRC	Storm source description	String	"Tropical Cyclone"
TCDVLP	Development status (e.g., "Tropical Depression")	String	"Tropical Storm"

3. Data Preprocessing

3.1 GIS Nature & Challenges

- CRS Handling: All spatial layers were checked and, if necessary, reprojected to a consistent coordinate reference system (CRS), specifically using UTM zone 16N (EPSG:32616) for accurate spatial computation.
- Geometry Extraction: For lines and polygons, centroids or start points were extracted to obtain latitude and longitude for modeling.

3.2 Python Libraries Used

- geopandas: For reading, manipulating, and reprojecting GIS shapefiles.
- tensorflow/keras: For deep learning models (RNN, LSTM).
- pandas: For tabular data manipulation and cleaning.
- numpy: For numerical operations and array handling.
- matplotlib: For plotting and visualizing spatial and prediction results.

- scikit-learn: For machine learning models and metrics.

3.3 Visualization Preprocessing

- Map Preparation: using GIS system - latitude & longitude

4. Algorithm(s) Used

4.1 Random Forest (RF)

- Why Chosen: Robust to nonlinear relationships, handles tabular and engineered features well, and is effective with moderate-sized datasets.
- Usage: Trained separately for each layer using available features (e.g., [LAT, LON, FCSTPRD] or [LAT, LON, MAXWIND, MSLP]).

4.2 Recurrent Neural Network (RNN)

- Why Chosen: Designed for sequence modeling, captures temporal dependencies in the hurricane's movement.
- Usage: Uses a sequence of previous time steps to predict the next [LAT, LON] position.

4.3 Long Short-Term Memory (LSTM)

- Why Chosen: An advanced type of RNN, LSTM is capable of learning longer-term dependencies and is well-suited for time series with complex temporal patterns.
- Usage: Similar to RNN, but with improved performance for longer sequences.

5. Quantitative Results (from Notebook Output/PDF)

5.1 Modelling GIS Data

RF significantly outperforms LSTM, RNN especially in linear & polygons layers

Mean Absolute Error:

Model/Layer	Linear	Polygon	Point	Warning line
RF	4.3	3.1	0.5	1.8
RNN	~40	40	5	3
LSTM	~45	47	6	3

Interpretation

- Low MAE: 1 degree (unit MAE) represent ~111km in RF

- Layer Differences: The _5day_pts layer (with richer meteorological features) yielded the best accuracy.
- Model Comparison: LSTM and RNN models leveraged temporal information, but RF performed surprisingly well, likely due to effective feature engineering & relatively short, regular sequences.

5.2 Forecasting IDALIA - 1 step ahead
Learnt a model based on previous hurricane data (sourced from NHC [dataset]) & applied 1-step ahead forecast to each point in IDALIA data

RF significantly outperforms LSTM, RNN especially in linear & polygons layers but performs worse in point & warning line
Mean Absolute Error:

Model/Layer	Linear	Polygon	Point	Warning line
RF	5.5	4.1	6.7	3.2
RNN	32	36	6.2	3.2
LSTM	41	34	6.1	3.1

Interpretation

- Low MAE: 1 degree (unit MAE) ~ 111km, in RF & point, warning line
- Layer Differences: The _5day_pts layer (with richer meteorological features) yielded the best accuracy, while purely geometric layers (_5day_lin, _ww_wwlin) had slightly higher errors.
- Model Comparison: LSTM and RNN models leveraged temporal information & performed better in point & warning line (more features), but RFperformed surprisingly well, likely due to effective feature engineering & the relatively short, regular sequences.

6. Graphical Results (from Notebook Output/PDF)

Note: The following descriptions refer to the plots generated in the Colab notebook.

- Actual vs Predicted Paths: For each layer, the actual (blue) and predicted (red) hurricane tracks are plotted on a longitude-latitude map.
- Directionality: Arrows indicate the direction of movement, and each vertex is labeled with its day number.
- MAE Displayed: Forecasted plot title includes the Mean Absolute Error

Example Plot Features:

- Blue circles and day labels: Actual hurricane path.
- Red crosses and day labels: Model-predicted path.
- Arrows: Show the direction of movement for both actual and predicted tracks.
- Grid and axes: Longitude (x-axis), Latitude (y-axis).

7. Conclusion

RF significantly outperforms LSTM, RNN especially in linear & polygons layers

8. Acknowledgement

We would like to express our sincere gratitude to Prof. Faiz Hamid for allowing us to work on such an interesting topic & for his guidance, support, and encouragement throughout this course..

This report summarizes the end-to-end workflow for hurricane path prediction, including data preprocessing, model selection and training, results analysis. All code, results, and plots are reproducible from the provided Colab notebook.