Advancing Weather Forecasting: A Comparative Analysis of Long Short-Term Memory and Support Vector Machines

Syed Muhammad Mohsin

Muhammad.Mohsin.2@city.ac.uk

Abstract

This research delves into the complex domain of weather forecasting, focused on the two fundamental meteorological parameters i- e rainfall and temperature. Leveraging the capabilities of two machine learning and deep learning models we have conducted a comparative study between Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) in forecasting weather. Through rigorous experimentation we provide valuable insights and enhance these models reliability to help mitigate weather related risks and optimize resource allocation.

1. Introduction:

Weather forecasting stands as a cornerstone in facilitating informed decisions across multiple sectors ranging from energy production, agriculture sector to disaster management . Among the abundance of variables temperature and rainfall hold the most significance because of their far-reaching implications in irrigation and water management, infrastructure planning and in general day to day human well-being. Historically. weather forecasts have relied on conventional methods and are dumb founded in capturing complex weather patterns and its non-linear relationships.

In this study, we embark on a comprehensive investigation into the application of LSTM and SVM to predict weather efficiently and accurately. We critically analyse the dataset and variables involved and their importance in developing an optimal model for rainfall and temperature predictions. Designed the architecture of SVM and LSTM to enhance the models capabilities and assessed their performance and precision over spatiotemporal projections. As a result, suggesting a conclusion which sheds light on their potential synergies and complementarities.

2. Models Opted: General Comparison

2.1. Long Short-Term Memory (LSTM):

Long Short-Term Memory (LSTM) is an advanced form of recurrent neural networks (RNNs). It is designed to capture long dependencies in sequential data. LSTM has input layer, hidden layers and an output layer but unlike traditional RNNs it incorporates memory cells and gates to hold information over long time intervals. LSTM retains or discards information through its gated units allowing it to mitigate the vanishing gradient problem usually encountered in traditional RNNs.

2.2. Support Vector Machine (SVM):

Support Vector Machines (SVM) is a supervised learning algorithm mostly used for classification but can be used in regression tasks. It works by creating an optimal hyperplane that differentiates between continuous outcomes and classifications. The hyperplane is constructed by maximizing the margins between the regression tasks and classes. This approach allow SVMs to do well with unseen and noisy data.

LSTMs networks are great with sequential data making them suitable for time-series predictions, while SVMs are proficient in handling high dimensional data including weather forecasting.

3. Dataset:

The dataset offers a rich source of information for understanding and analysing weather patterns. The dataset is for the region of Ireland. It has 662,345 samples(rows) and 27 variables (columns).

Among the key predictors in the dataset are rainfall (rain) and temperature (temp), which play a pivotal role in climate analysis. Rainfall is measured in millimetres while air temperate is recorded in Celsius °C.

In addition to rainfall ,temperature and the attributes shown in table() below the dataset have geographical attributes such as station ID (st_id), station Name (st_name) , station height (st_height), latitude (st_lat), longitude (st_long) and county.

No	Variable	Description	Type
1	date	Date and Time of observation	datetime64[ns]
2	ind	Encoded Rainfall Indicators	int64
3	rain	Precipitation Amount, mm	float64
4	ind.1	Encoded Temperature Indicators	int64
5	temp	Air Temperature, °C	float64
6	ind.2	Encoded Wet Bulb Indicators	int64
7	wetb	Wet Bulb Air Temperature, °C	float64
8	dewpt	Dew Point Air Temperature, °C	float64
9	vappr	Vapour Pressure, hPa	float64
10	rhum	Relative Humidity, %	float64
11	msl	Mean Sea Level Pressure, hPa	float64
12	ind.3	Encoded Wind Speed Indicators	int64
13	wdsp	Mean Hourly Wind Speed, knot	int64
14	ind.4	Encoded Wind Direction Indicators	int64
15	wddir	Predominant Hourly wind Direction, degree	float64
16	ww	Synop Code Present Weather	int64
17	w	Synop Code Past Weather	int64
18	sun	Sunshine duration, hours	float64
19	vis	Visibility, m	int64
20	clht	Cloud Ceiling Height (if none value is 999), 100s of feet	int64
21	clamt	Cloud Amount, okta	int64

Fig.1 Variables of dataset

A detailed analysis of dataset's summary reveals important insights to each attribute. The mean temperature is around 9.92 °C and std of 4.83 °C reflecting fluctuations in thermal conditions. Climate is mild with higher rainfalls but lack of extreme temperatures. January and February are coldest months with mean falling to 4 °C but July and August are warmest with mean going up to 16 °C. The average rainfall is approximately 0.12 mm with std of 0.50 mm which indicates varying precipitation levels. Wettest months are December and January. The mean wet days in a year in coastal regions are about 151 days.

Attributes	count	mean	std	min	25%	50%	75%	max
rain	662345	0.12438	0.497221	0	0	0	0	38.4
temp	662345	9.922873	4.830016	-15.4	6.6	10	13.4	31.5
wetb	662345	8.596542	4.525487	-49.9	5.5	8.9	11.9	22
dewpt	662345	7.185166	4.513444	-16.4	3.9	7.4	10.6	19.9
vappr	662345	10.57009	3.175379	1.7	8.1	10.3	12.8	23.2
rhum	662345	84.14293	12.11484	20	77	87	94	100
msl	662345	1013.246	12.5294	948.2	1005.5	1014.4	1021.9	1050
wdsp	662345	9.918461	5.575393	0	6	9	13	61
wddir	662345	200.9745	84.90956	0	140	220	260	360
ww	662345	17.18909	23.64326	0	2	2	25	97
w	662345	33.52028	28.70327	0	11	11	62	99
sun	662345	0.167173	0.324983	0	0	0	0.1	1
vis	662345	26918.63	15137.28	5	17000	25000	40000	75000
clht	662345	270.4222	401.6696	0	20	41	250	999
clamt	662345	5.74639	2.32382	0	4	7	7	9
st_id	662345	2477.993	1668.161	518	532	2375	3904	4935
st_height	662345	86.5596	63.48279	9	20	71	155	201
st_lat	662345	53.04225	0.842819	51.84722	52.69028	53.30556	53.42778	55.37194
st_long	662345	-7.96417	1.358236	-10.2408	-8.91806	-8.48611	-6.43889	-6.24083

Fig.2 Descriptive Statistics

4. Exploratory Data Analysis (EDA):

In EDA first we examined the correlation matrix to understand the relationships between attributes.

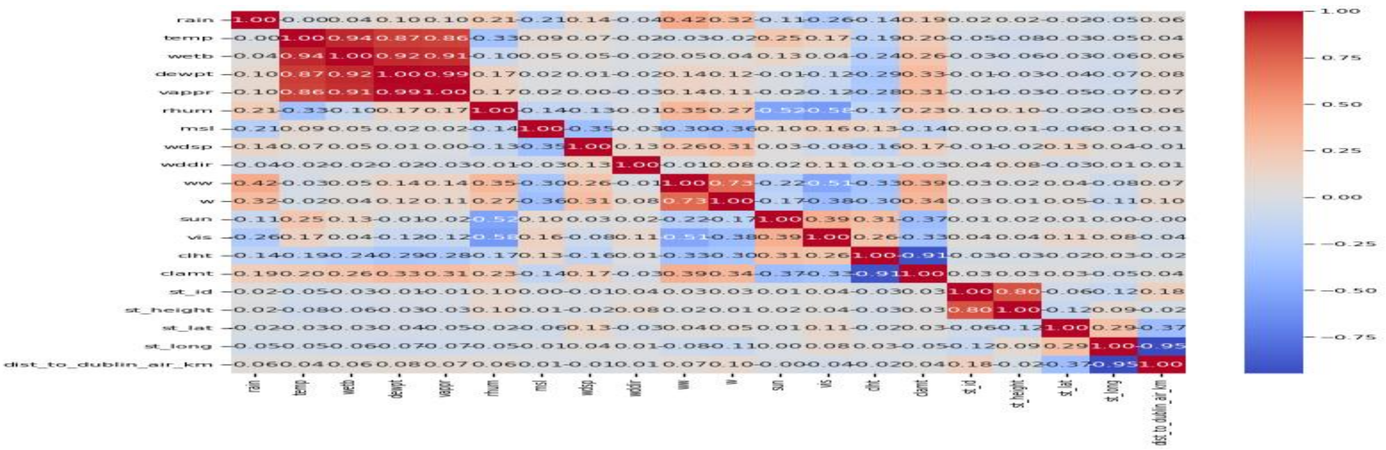


Fig.3 Correlation matrix

Variables like Rainfall (rain) and other weather indicators like humidity and wind speed showed positive correlations with each other. This exhibits abundance of rainfall with increased humidity and wind speed levels. On the other hand ,humidity is negatively correlated with temperature (temp) suggesting lower humidity levels with higher temperatures. Realization of significant and non-significant variables can also be deduced from the correlation matrix. For example temperature and rainfall are not impacted by cloud ceiling levels and wind direction.

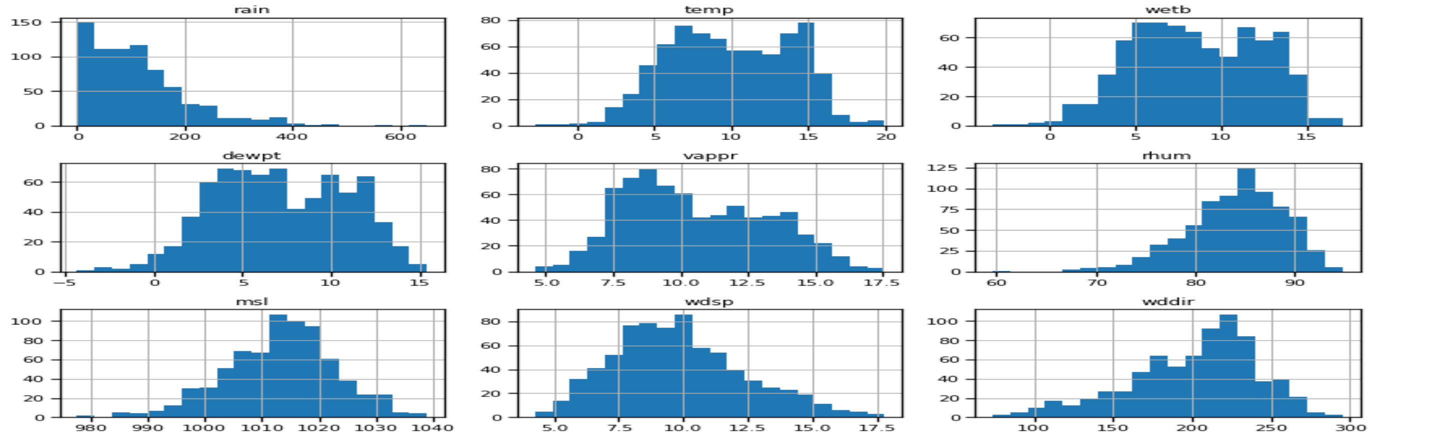


Fig.4 Distribution Graphs of variables

Some of the key findings are obtained after looking at the distribution of variables. Rainfall is the only skewed right that means positively skewed which depicts that high number of rainfall of low amounts and fewer number of high rainfall instances. Others variables are mostly normally distributed. The temperature, wet bulb indicator and dew indicator shows two peaks which means two seasons of variations possibly winter and summer. Looking at the distribution allowed us to built the concept that our data shows seasonal aspect along with just the trends and patterns in the weather.

5. Methodology:

The dataset underwent quality checks, revealing no missing values. Also highlighting unique time stamps which were around 123503 values. After removing nulls and interpolating them with means 'weekly data' was created by aggregating rainfall measurements and averaging meteorological attributes. This finalized dataset provides a clear view of weekly weather patterns, facilitating accurate analysis and modelling for forecasting applications. The data normalization process using Min-MaxScaler rescales the rainfall data to a range of -1 to 1, ensuring consistency and aiding model convergence. Following normalization, the data is converted into PyTorch tensors for compatibility with neural network architectures. Subsequently, sequence of 5 steps were created to ensure 5 consecutive captures revealed. This is useful for sequential modelling

Finally the data set is split into 80% training – 20% testing sets. Minimum date in the data is '2007-12-31 02:00:00' and maximum date is '2022-02-01 00:00:00' which gives around 736 weeks in total for the aggregated data. Given 736 (100%) rows , 589 (80%) rows were used for training and 147 (20%) rows were used for testing and evaluation of the model.

5.1. Model Training:

Both the SVM and LSTM models were trained using the training data. During training hyperparameter of the models were adjusted to reach a certain level of precision in forecast.

5.1.1. Architecture of LSTM (Model training):

The LSTM (Long Short-Term Memory) model consists of an input layer, an LSTM layer, and an output layer. For the training purpose epochs for LSTM are set to 50 to get the most accurate results. However this is just to deduce which set of hyperparameters will yield great results. Best set of hyper parameters based on lowest validation losses and minimum training times were selected and inserted to create the best model to represent LSTM. The input size is set to 1, indicating the number of features per time step. The hidden layer sizes are set to either 5 or 50 depending on the model training times and validation losses, allowing the model to capture temporal dependencies in the data. Also, different learning rates of either 0.001 or 0.01 were set to get full trend in the data. Finally, I have set a single output layer which maps the LSTM desired output. The idea is to create the architecture in this way is to get the temporal trends in a more optimized way.I have used Adam (Adaptive moment Estimation) as an optimization technique in LSTM. Adam used momentum because it handles acceleration of gradients and dampens oscillations furthermore it corrects bias to mitigate the effects of initialization of moments to 0. Using Adam was very effective for noisy data.

5.1.2. Architecture of SVM (Model training):

The SVM (Support Vector Machine) model operates on a different principle compared to LSTM since it doesn't need weights initialising and concepts of hidden layers and neurons.

SVM extension SVR (Support Vector Regressor) model was used to predict rainfall. Defining the kernel function for SVMs takes the utmost priority so I have used 'rbf' and 'linear' kernels to get the understanding of the data type. Different Regularization parameters were also used minimize the errors in regressions. Finally , Grid search with cross validations set to 3 folds was implemented to check the robustness and reliability of the model. Hyperparameters for SVM shown in Fig. 5.

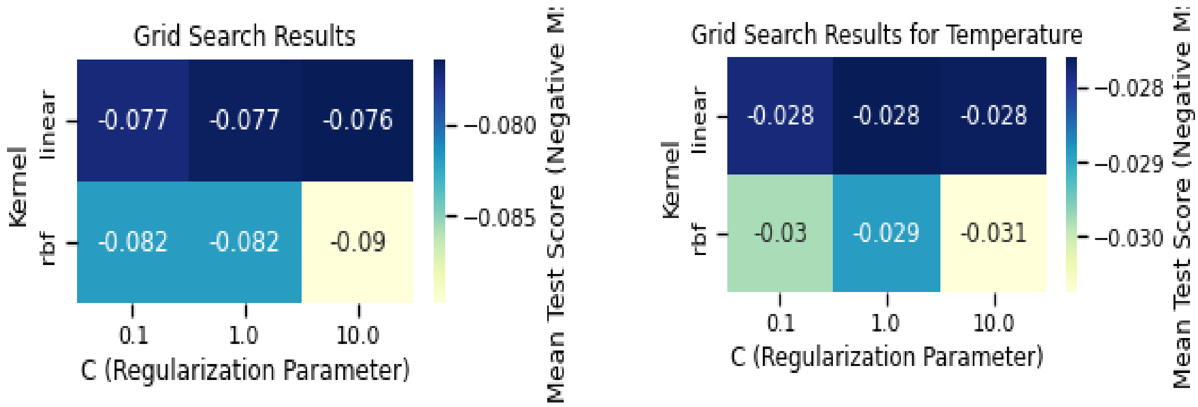


Fig.5 Hyperparameters SVM Matrix vs Mean Squared Error

In the end both of the models forecasts the temperature and rainfall which were then evaluated through statistical methods of Mean Squared Error (MSE) ,Mean Absolute Error (MAE) and residuals test.

6. Results, Findings and Evaluations:

The table in the picture below Fig.6 shows the different iterations of the hyperparameters and the selection of the best(yellow) and worst(red) hyperparameters for both the models. Epochs for LSTM are set to 50 for identifying the best parameters. The test is entirely on the basis of the fact that which set of hyperparameters results in the optimal performance of the models in cases of Rainfall and Temperature. In this grid search process we have realized that for SVM, the optimal model is an RBF kernel with a regularization parameter (C) of 0.1 and gamma scale. This can be seen that it has the lowest validation losses and a very effective training time. SVM configurations indicate a superior predictive accuracy on the unseen data. On the other hand, LSTM exhibits a higher validation losses and training times where values goes beyond 1 seconds for each iteration meaning that it is not efficient to generalize the unseen data as much as SVM.

	Hyperparameters			Temperature		Rain	
	Kernel	C	Gamma	Validation Loss	Training Time	Validation Loss	Training Time
SVM	rbf	0.1	scale	-0.0310	0.0040	-0.0845	0.0067
	rbf	0.1	auto	-0.0278	0.0040	-0.0773	0.0053
	rbf	1	scale	-0.0286	0.0027	-0.0794	0.0066
	rbf	1	auto	-0.0278	0.0053	-0.0773	0.0040
	rbf	10	scale	-0.0308	0.0080	-0.0876	0.0093
	rbf	10	auto	-0.0276	0.0080	-0.0766	0.0067
	linear	0.1	scale	-0.0270	0.0053	-0.0763	0.0053
	linear	0.1	auto	-0.0276	0.0093	-0.0766	0.0080
	linear	1	scale	-0.0335	0.0147	-0.1007	0.0187
	linear	1	auto	-0.0276	0.0146	-0.0764	0.0133
	linear	10	scale	-0.0280	0.0093	-0.0789	0.0080
	linear	10	auto	-0.0276	0.0147	-0.0764	0.0147
	Learning Rate		Hidden Size	Validation Loss	Training Time	Validation Loss	Training Time
LSTM	0.001		10	0.0334	1.2838	0.0334	1.1610
	0.001		50	0.0333	1.2569	0.0332	1.0763
	0.01		10	0.0346	1.1722	0.0344	1.2534
	0.01		50	0.0345	1.1404	0.0342	1.1363

Fig.6 Model Selection SVM vs LSTM (Validation Losses and Training Times)

6.1. Choosing the best Model:

Though with the hyperparameter results we can analyse that both of these models performed relatively close with respect to predictions of rainfall. However, both models exhibit some degree of error, the LSTM outperforms SVM in predicting rainfall. LSTM demonstrates some powerful traits to capture underlying patterns and trends in the rainfall .LSTM model shows more consistency and stability compared to SVM in terms of predictions overall.

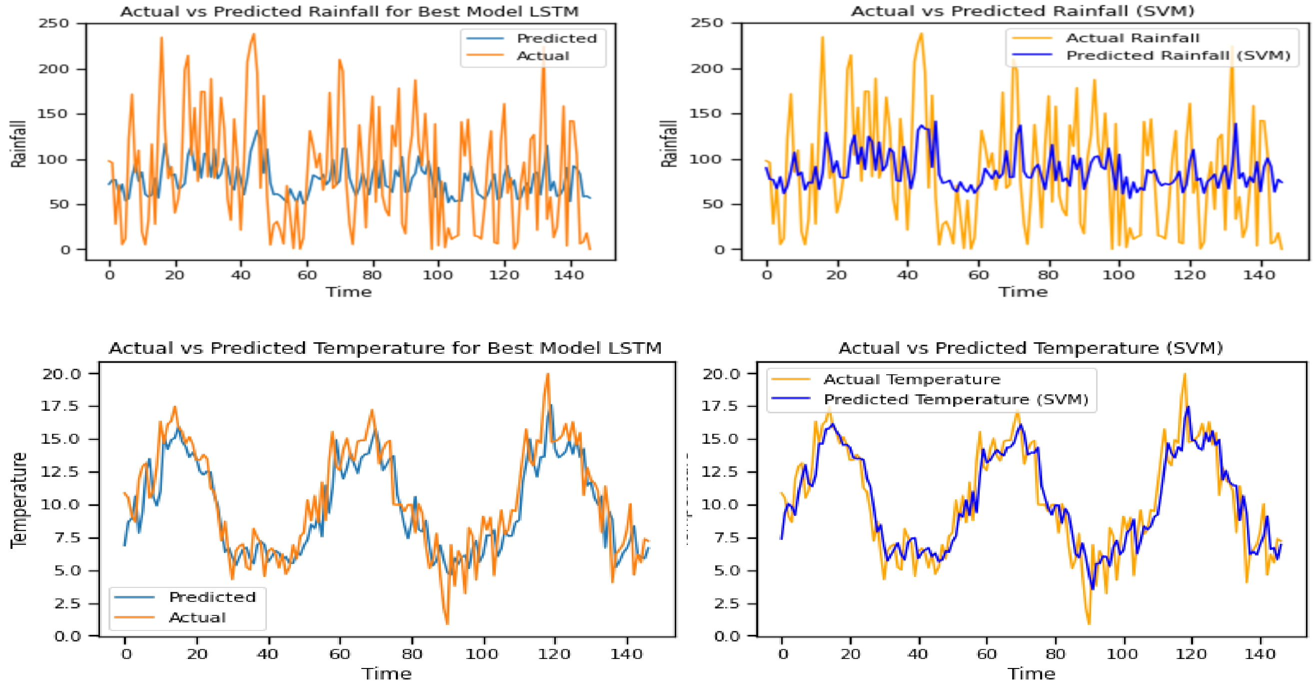


Fig.7 Rainfall and Temperature Predictions

Both Models made predictions close to actual predictions but with data points more close to each other a quick change in temperature values leads to errors in forecasts. LSTM model closely tracked the actual trends in the temperature which reflects its ability to capture temporal patterns more effectively. In contrast, SVM exhibits higher errors indicating less precision and more biasness. This can be seen from the graphs above that SVM tracks the actual predictions without allowing itself the room to grasp the change in temperatures. It is less robust to change. The ability to generate accurate results and consistent forecasts actually underscores SVM's ability to be applicable in the real world applications.

6.2. MSE/MAE AND RESIDUALS:

Comparing the performance of the models based on temperature we see that SVM has a lower MSE of 3.51 instead of 3.63 for the LSTM shown (Fig.8). This is because SVM model might handle the outliers better in the data and penalizes large errors more heavily. In terms of MSE alone LSTM is a poor model. However, MAE in general for the LSTM is lower which shows precision and accuracy of the models. The MAE of LSTM is 50.75 and for SVM is 51.57 which shows more close predictions to the average for LSTM. In case of weather forecasts where consistent results matter LSTM proves to be a better option than SVM. Though the MAE for temperature is 1.45 for SVM compared to 1.51 for LSTM shows the consistency of LSTM over SVM in general even though SVM shows biasness shown in Fig 9.

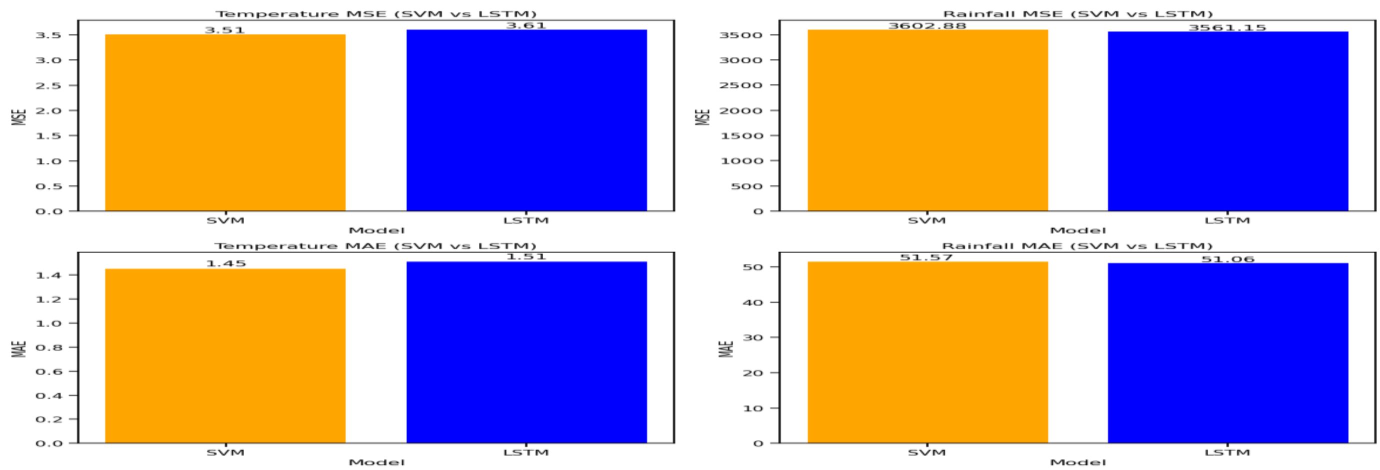


Fig.8 Mean Squared Error (MSE) & Mean Absolute Error (MAE) – LSTM vs SVM

The residual plot below also highlight that rainfall predictions for both SVM and LSTM are mostly based around the 0. But SVM have more data points closer to 0 means it generally deals with outliers better. The residual plots for temperature however , expose the SVM Models indicating that almost all the data points fall below 0. This gives a clear picture stating that SVM is biased for temperature predictions. This could be due to the inherent limitation in the SVM algorithm for the given dataset and features. SVM prefers more complex features to ensure robustness in the model.

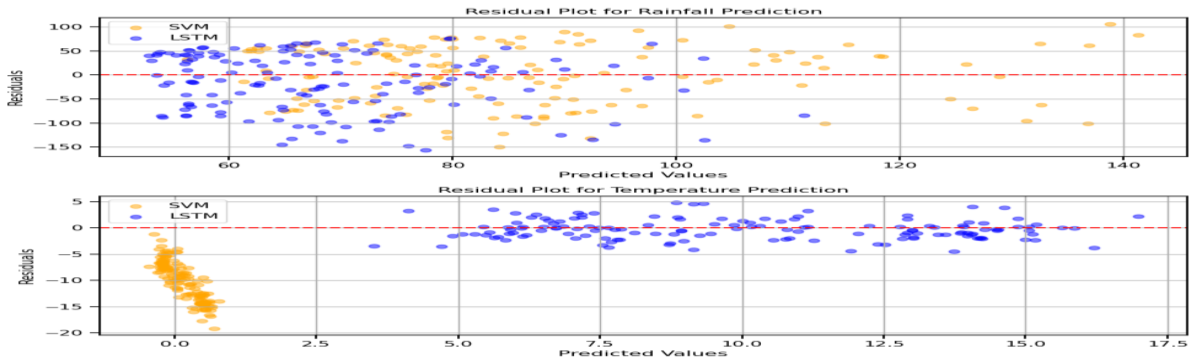


Fig.9 Residual Plot – SVM vs LSTM

7. Conclusion:

Our comparative analysis between Long Short-Term Memory (LSTM) and Support Vector Machines (SVM) models for weather forecasting highlights the strengths and weaknesses of each approach. After looking at MSE and MAE in general, LSTM demonstrates superior performance in capturing temporal dependencies but loses to SVM in terms of handling outliers and rainfall predictions. Considering also the fact that LSTM takes longer times to train , greater validation loses, and is also it is not easy to implement unless fine-tuned. To conclude, weather conditions and dataset necessitates a careful consideration of forecast requirements for an optimal model solution.

8. Future Work:

It is wise to conduct a more rigorous approach in using hyper parameters for both LSTM and SVM models in the future. Consider learning rates below 0.01 for LSTM to get more model convergence for example in rainfall scenario. Bi-directional LSTMs could also be beneficial because understanding past and future at the same time is crucial in weather forecasts. Expanding the Regularization for SVM should also yield better results because it will remove the problems of overfitting which seems to be the case in temperature predictions. No basic model is sufficient so it is imperative to harness the full potential of these models for predicting weather forecasts and understand weather phenomena.

9. References:

[1] Billah, Mustain, Adnan, Md. Nasim, Akhond, Mostafijur Rahman, Ema, Romana Rahman, Hossain, Md. Alam and Md. Galib, Syed. "Rainfall prediction system for Bangladesh using long short-term memory" Open Computer Science, vol. 12, no. 1, 2022, pp. 323-331. https://doi.org/10.1515/comp-2022-0254.

[2] Nandakumar, S D, et al. “Artificial Neural Network for rainfall analysis using Deep Learning Techniques.” Journal of Physics: Conference Series, vol. 1964, no. 4, 1 July 2021, p. 042022, https://doi.org/10.1088/1742-6596/1964/4/042022.

[3] Golian, Saeed, et al. “Dynamical–statistical seasonal forecasts of winter and summer precipitation for the island of Ireland.” International Journal of Climatology, vol. 42, no. 11, 15 Feb. 2022, pp. 5714–5731, https://doi.org/10.1002/joc.7557.

[4] Abhishek, Kumar, et al. “A rainfall prediction model using artificial neural network.” 2012 IEEE Control and System Graduate Research Colloquium, July 2012, https://doi.org/10.1109/icsgrc.2012.6287140.

10. Appendix:

10.1. Glossary

1. Weather Forecasting: Process of predicting weather conditions from past data.

2. Long Short-Term Memory (LSTM): A type of recurrent neural network (RNN) architecture designed to capture long-term dependencies in sequential data.

3. Support Vector Machines (SVM): A supervised learning algorithm used for classification and regression tasks.

4. Exploratory Data Analysis (EDA): The process of examining and visualizing data

5. Grid Search: A method used to find the optimal combination of hyperparameters

6. Cross-Validation: Different combinations of the dataset

7. Mean Squared Error (MSE): A statistical measure of accuracy

8. Mean Absolute Error (MAE): A statistical measure of accuracy.

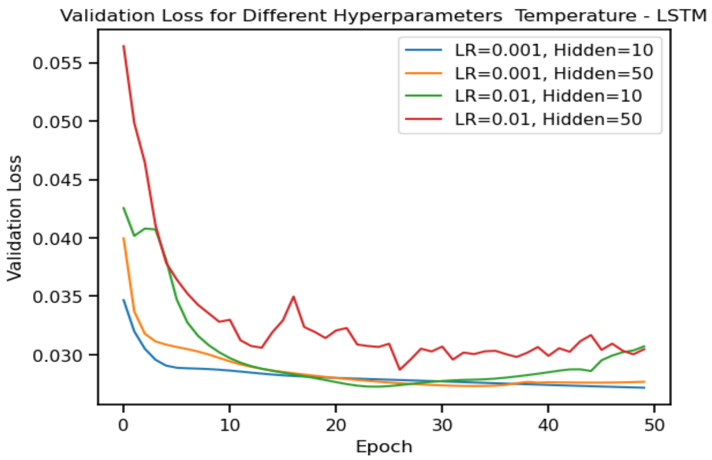
9. Residuals: The differences between observed and predicted values

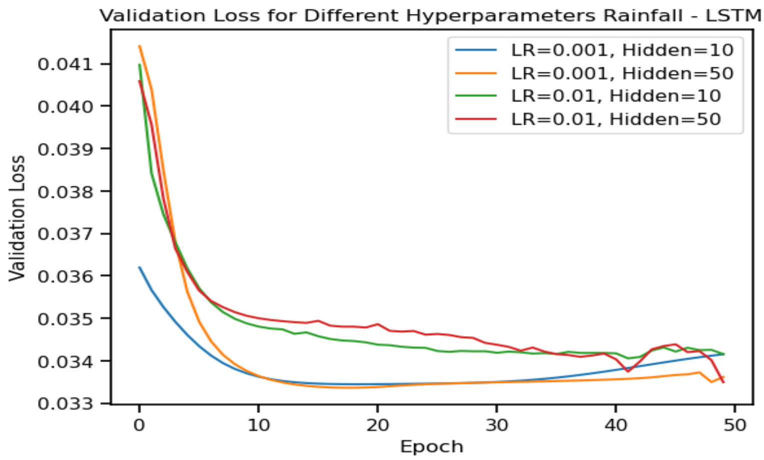
10. Hyperparameters: Parameters that are set prior to model training and affect the learning process

11. Optimization Techniques: Methods used to optimize model parameters during training

12. Robustness: The ability of a model to perform well on unseen data

10.2. Intermediatory Results:





This the validation loss graphs for the LSTM model of rain and temperature. In general they exhibit same trends and have low validation losses for most set of parameter sets. However, learning rate of 0.001 tend to have a lower validation losses compared to 0.1 in both the models. And hidden size increase to 50 results in a lower loss meaning better performance. This concludes that using a learning rate of 0.001 and hidden size 50 usually yields a better model performance.

10.3. Word Count:

Document	Word Count
Abstract	74
Introduction	155
Model Opted	167
Dataset	206
Exploratory Data Analysis	194
Methodology	585
Results,Findings & Evaluation	623
Conclusion	97
Future Work	111
Glossary & Intermedtiatory Results	232
Total	**2444**