



A Hybrid Approach to Sales Forecasting: Combining Deep Learning and Machine Learning

By

Othman Omar Othman Alamoudi

Midocean University

Under Supervision of

Prof. Dr. Yasser Abdelhamid

Abstract

Sales forecasting is important for businesses to thrive, which for the plan preparations for next year sales. Traditional sales forecasting often struggle due to many factors human mistakes in the calculation formula or delay for the plan preparations or any other reason. This research targeting the advantages of deep learning and machine learning algorithms, like LSTM, Random Forest, Naive bayes, XGBoost... and if there is any chance to develop a hybrid model that will improve sales predictions using historical sales data. The sales forecasting has a significant shift towards advanced methods, with machine learning and deep learning techniques experiencing an explosive increase in adoption. The popularity of these models has surged, as evidenced by a rise from 10 articles in 2013 to over 110 by 2023[1], this research aims to evaluate and to pick the best algorithm that will provide the best and efficiently results the sales forecasting. This research will help companies for their sales forecasting plans by using the power of AI.

Keywords: Deep Learning, Machine Learning, Hybrid Machine and Deep learning, Sales Forecasting, Retail.

Introduction

Sales forecasting is crucial element of supply chain management and it became very important specially with the products that

have short life cycle which is challenging for the companies due to many factors like fears of the loss due to slow movements of the products or incapability of storing those products.

Q1. What are the DL and ML models used for sales forecasting?

Q2. What metrics are used for model evaluation?

Q3. What are the benefits of using Hybrid Learning models for sales forecasting?

Q4. What are the challenges and limitations of using Hybrid Learning models for sales forecasting?

Problem Definition

In the most companies the sales forecasting is important to make the vision for the next year and also to prepare the plan for the logistics and supply chain processes which currently prepared using the traditional ways which often have inaccurate results which this study aims to improve the process to get the best result that we can.

Dataset

This section describes a collection of public datasets in different literature that are used in the field of brain damage diseases. Some literature has also used data collected from patients of various hospitals.

Dataset Name	Link	Biomarker	Size & Samples	Limitations
Amazon Dataset	Kaggle	CSV Dataset	4MB , 1466	
Mart Dataset	Kaggle	CSV Dataset	24 MB , 537577 Record	

Related Work

Sales Prediction Using Machine Learning (2024)

This paper evaluates the effectiveness of various machine learning models, such as linear regression, random forest, and Gradient Boosting algorithm for predicting future sales.

The researchers collected historical sales data, preprocessed it, selected important features, and trained the models. They evaluated the models using metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

Handling large datasets and ensuring the accuracy of predictions.

Integrating more data sources and using advanced deep learning architectures for better accuracy.

Predicting Future Sales: A Machine Learning Algorithm Showdown (2024)

This paper explores the use of machine learning models, including linear regression, random forest, XGBoost, and LSTM, for forecasting future sales based on historical data.

The researchers used the "Store Item Demand Forecasting" dataset, comprising five years of sales data for 50 items across ten stores. They employed regression techniques and evaluated the models using MAE, RMSE, and R-squared.

Handling vast data volumes and ensuring accurate predictions.

Incorporating more sophisticated models and techniques to improve prediction accuracy.

Approaching Sales Forecasting Using Recurrent Neural Networks and Transformers (2022)

This paper investigates the use of deep learning techniques, such as sequence-to-sequence models, for daily sales forecasts.

The researchers applied a sequence-to-sequence architecture to customer sales data with minimal data preprocessing. They also described a training trick to improve the model's time generalization. The proposed solution achieves a RMSLE of around 0.54, which is competitive with other more specific solutions to the problem proposed in the Kaggle competition. The model evaluated using RMSLE (Root Mean Squared Logarithmic Error), RMSWLE (Root Mean Squared Weighted Logarithmic Error), MALE (Mean Absolute Logarithmic Error).

Ensuring the model's generalization over time and handling minimal data preprocessing.

Implementing more advanced deep learning techniques and incorporating additional data sources.

Retail Sales Forecasting Using Deep Learning: Systematic Literature Review (2022)

This paper reviews various deep learning models for retail sales forecasting, comparing their benefits and limitations with other machine learning and linear models.

The researchers conducted a systematic literature review, analyzing different deep learning frameworks and evaluation metrics used in the reviewed literature.

A Dynamic Ensemble Selection Method for Bank Telemarketing Sales Prediction (2022)

This paper presents a dynamic ensemble selection method for predicting bank telemarketing sales.

- Methodology: The researchers evaluated the performance of different machine

learning models and proposed an ensemble approach to improve prediction accuracy. The research Incorporating more diverse models in the ensemble and using advanced techniques to improve prediction accuracy.

Related Work Summary

References	Datatype	Training Model	Accuracy	Dataset
Sales Prediction Using Machine Learning (2024)	Historical sales data for Efashion stores	Linear regression, Random forest, Gradient Boosting algorithm	Mean Absolute Error (MAE), Root Mean Squared Error (RMSE)	Collected historical sales data
Predicting Future Sales: A Machine Learning Algorithm Showdown (2024)	Store item demand forecasting	Linear regression, Random forest, XGBoost, LSTM	Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-squared	Five years of sales data for 50 items across ten stores It contains a total of 913,000 rows, with each row representing sales data for a specific date, store and item
Approaching Sales Forecasting Using Recurrent Neural Networks and Transformers (2022)	Customer sales data	Sequence-to-sequence models (e.g., RNNs, Transformers)	Not specified	Minimal data preprocessing
Retail Sales Forecasting Using Deep Learning: Systematic Literature Review (2022)	Various literature data	Various deep learning models	Not specified	Systematic literature review
A Dynamic Ensemble Selection Method for Bank Telemarketing Sales Prediction (2022)	Bank telemarketing data	Ensemble method combining different machine learning models	Not specified	Evaluated different machine learning models

Research Objective

The objective of this study is to develop a Hybrid-based Regression model to predict the sales forecast using the machine and deep learning techniques to improve the accuracy of the sales forecasting. The system will be designed to determine the severity of the diseases using classification techniques.

- Selecting and implementing proper algorithms for sales forecast prediction.
- Evaluating the selected algorithms based on metrics such as accuracy, precision, and recall.
- Modifying the selected algorithms or proposing and designing new algorithms to overcome their drawbacks.
- Compare the obtained result with the other results obtained before for brain aging disorders.

References

- [1] Systematic Mapping Study of Sales Forecasting: Methods, Trends, and Future Directions by Systematic Mapping Study of Sales Forecasting: Methods, Trends, and Future Directions. 1 July 2024 by Hamid Ahaggach et al.
- [2] Sales Prediction Using Machine Learning (2024)
Hitesh S.M, Yukthi A, Ramya B.N. "Sales Prediction Using Machine Learning Techniques." International Journal of Novel Research and Development (IJNRD), 2024.
- [3] Predicting Future Sales: A Machine Learning Algorithm Showdown (2024)
Manal Loukili, Fayçal Messaoudi, Mohammed El Ghazi, Hanane Azirar. "Predicting Future Sales: A Machine Learning Algorithm Showdown." Lecture Notes in Networks and Systems (LNNS), Springer, 2024.
- [4] Approaching Sales Forecasting Using Recurrent Neural Networks and Transformers (2022)
Iván Vallés-Pérez, Emilio Soria-Olivas, Marcelino Martínez-Sober, Antonio J. Serrano-López, Juan Gómez-Sanchís, Fernando Mateo. "Approaching Sales Forecasting Using Recurrent Neural Networks and Transformers." arXiv, 2022.
- [5] Retail Sales Forecasting Using Deep Learning: Systematic Literature Review (2022)
Linda Eglite, Ilze Birzniece. "Retail Sales Forecasting Using Deep Learning: Systematic Literature Review." Complex Systems Informatics and Modeling Quarterly, 2022.
- [6] A Dynamic Ensemble Selection Method for Bank Telemarketing Sales Prediction (2022)
Y. Feng, Y. Yin, D. Wang, L. Dhamotharan. "A Dynamic Ensemble Selection Method for Bank Telemarketing Sales Prediction." Journal of Business Research, 2022.

A Hybrid Approach To Sales Forecasting

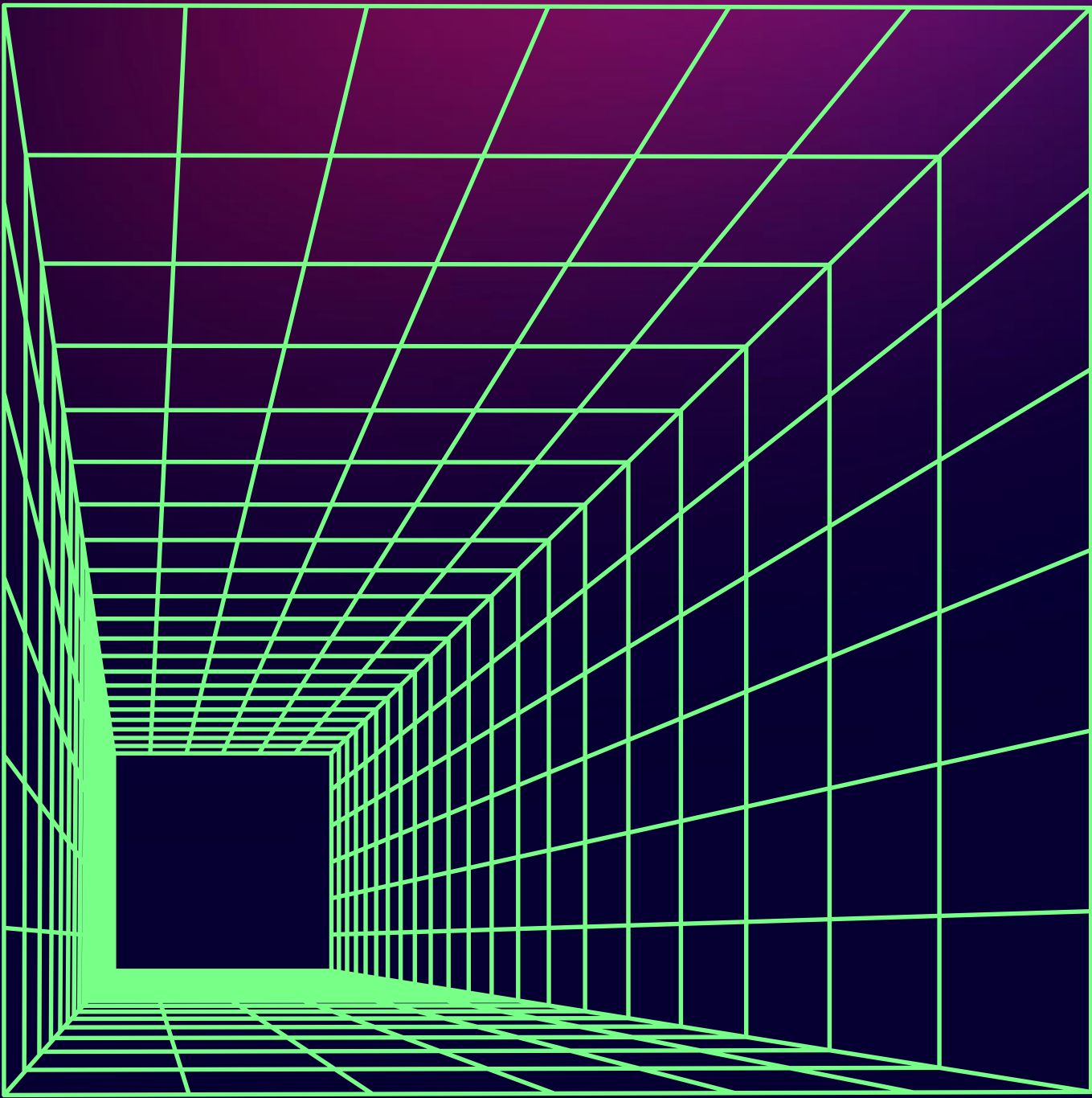
Othman Omar Othman Alamoudi
Midocean University

Prof. Dr. Yasser
Abdelhamid

Introduction

Traditional Sales forecasting

Keywords: Deep Learning, Machine Learning, Hybrid Machine and Deep Learning, Sales Forecasting.



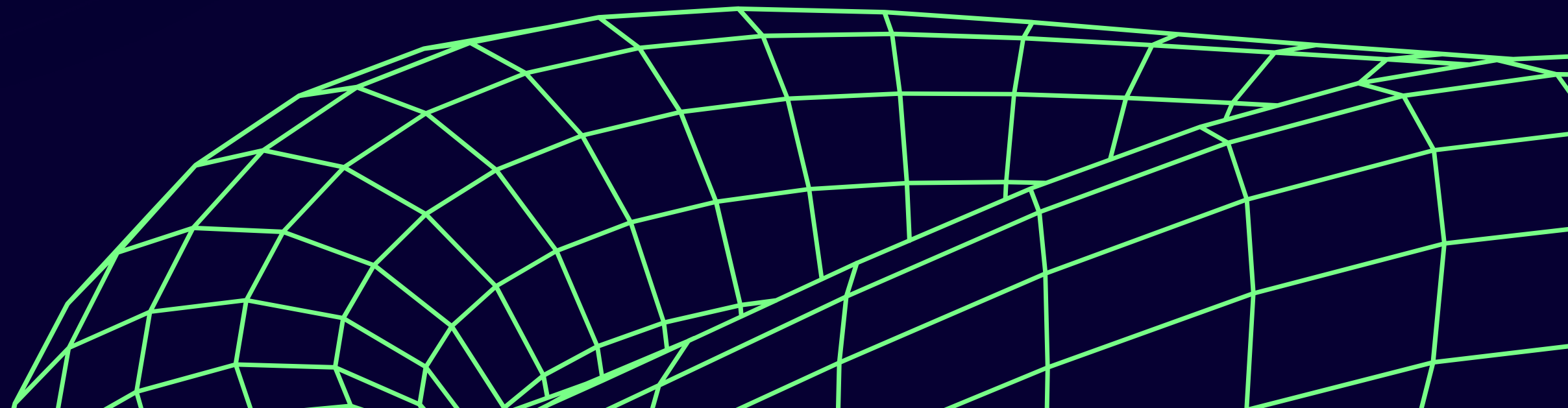
CONTENT

1 Problem Definition

2 Previous Related Work

3 Methodology Used

4 Evaluations Metrics

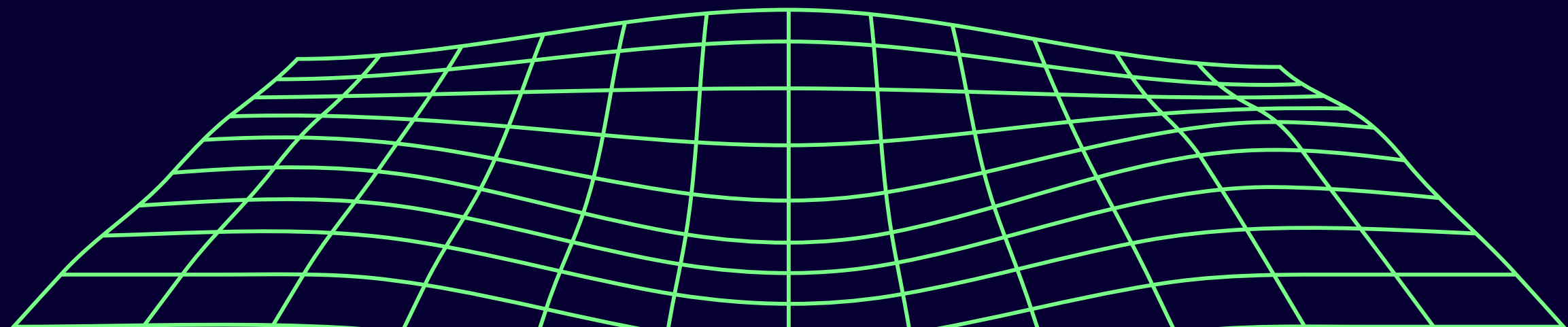


Problem Definition

Accuracy

Data Quality

Dynamic Market Conditions



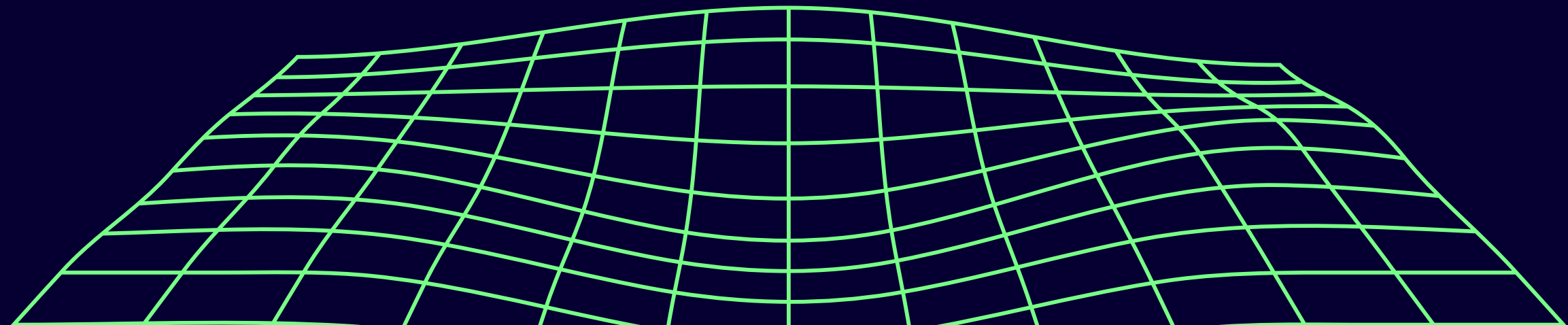
Research Questions

Q1. What are the DL and ML models used for sales forecasting?

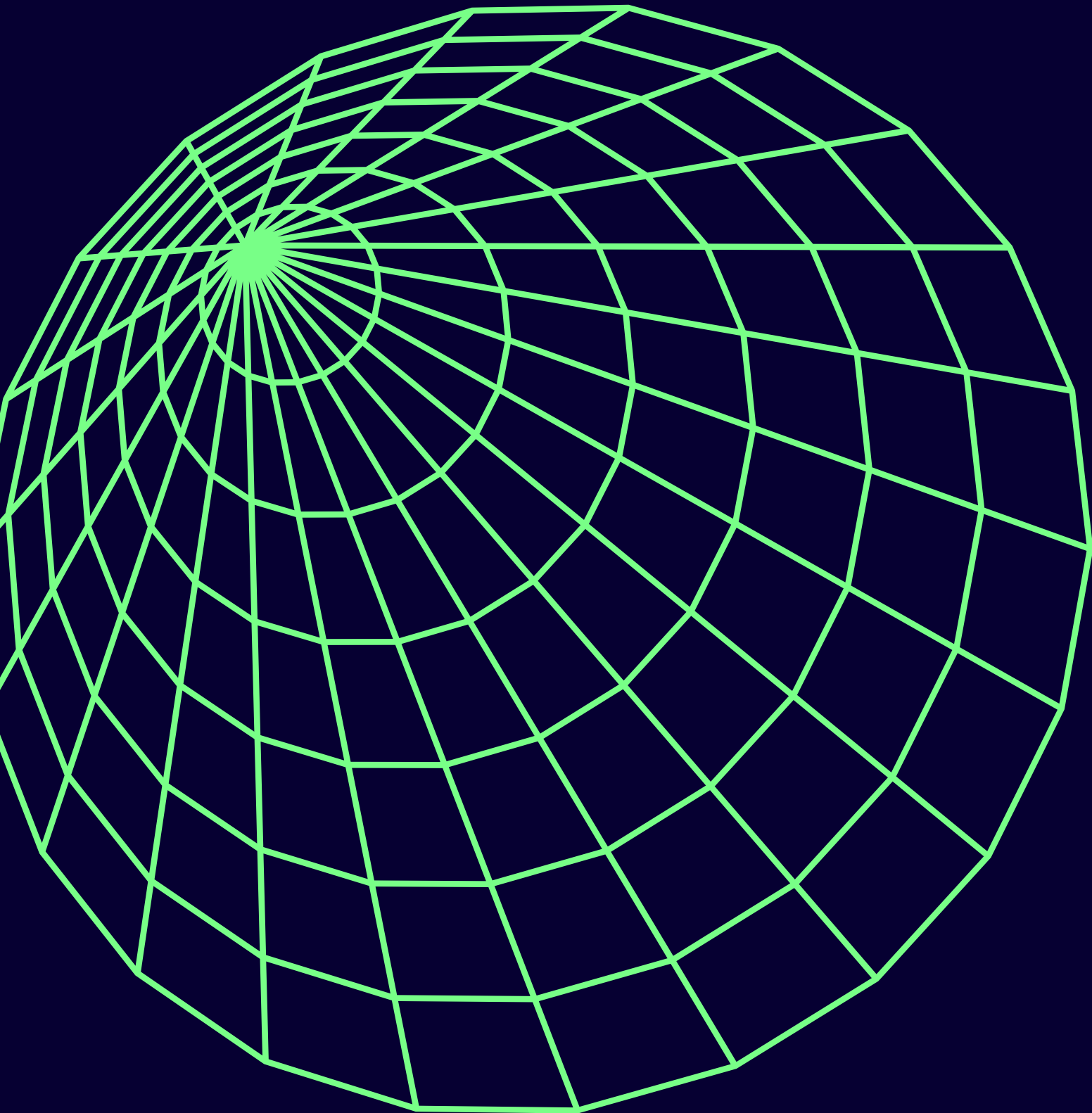
Q2. What metrics are used for model evaluation?

Q3. What are the benefits of using Hybrid Learning models for sales forecasting?

Q4. What are the challenges and limitations of using Hybrid Learning models for sales forecasting?

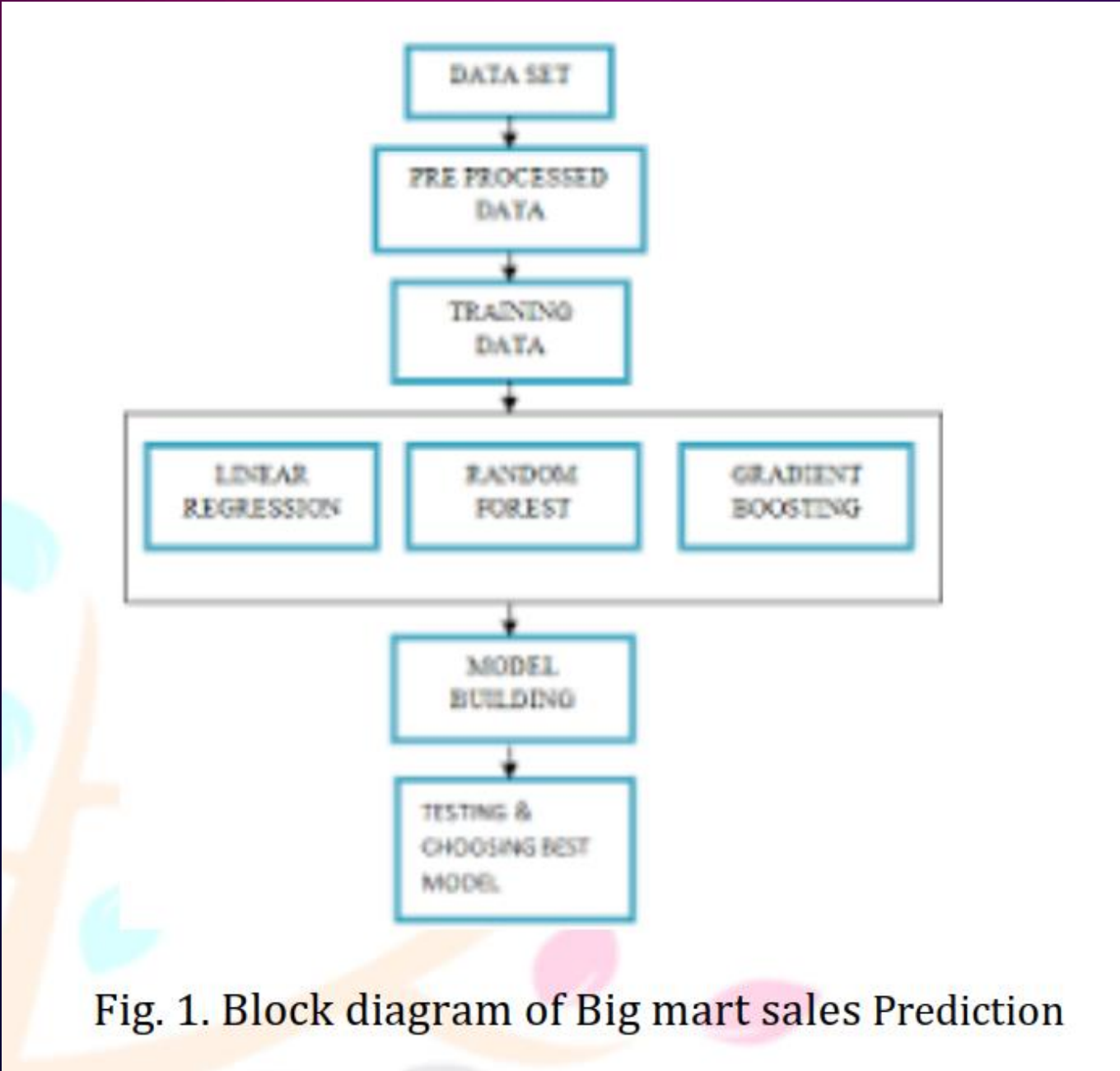


Machine Learning



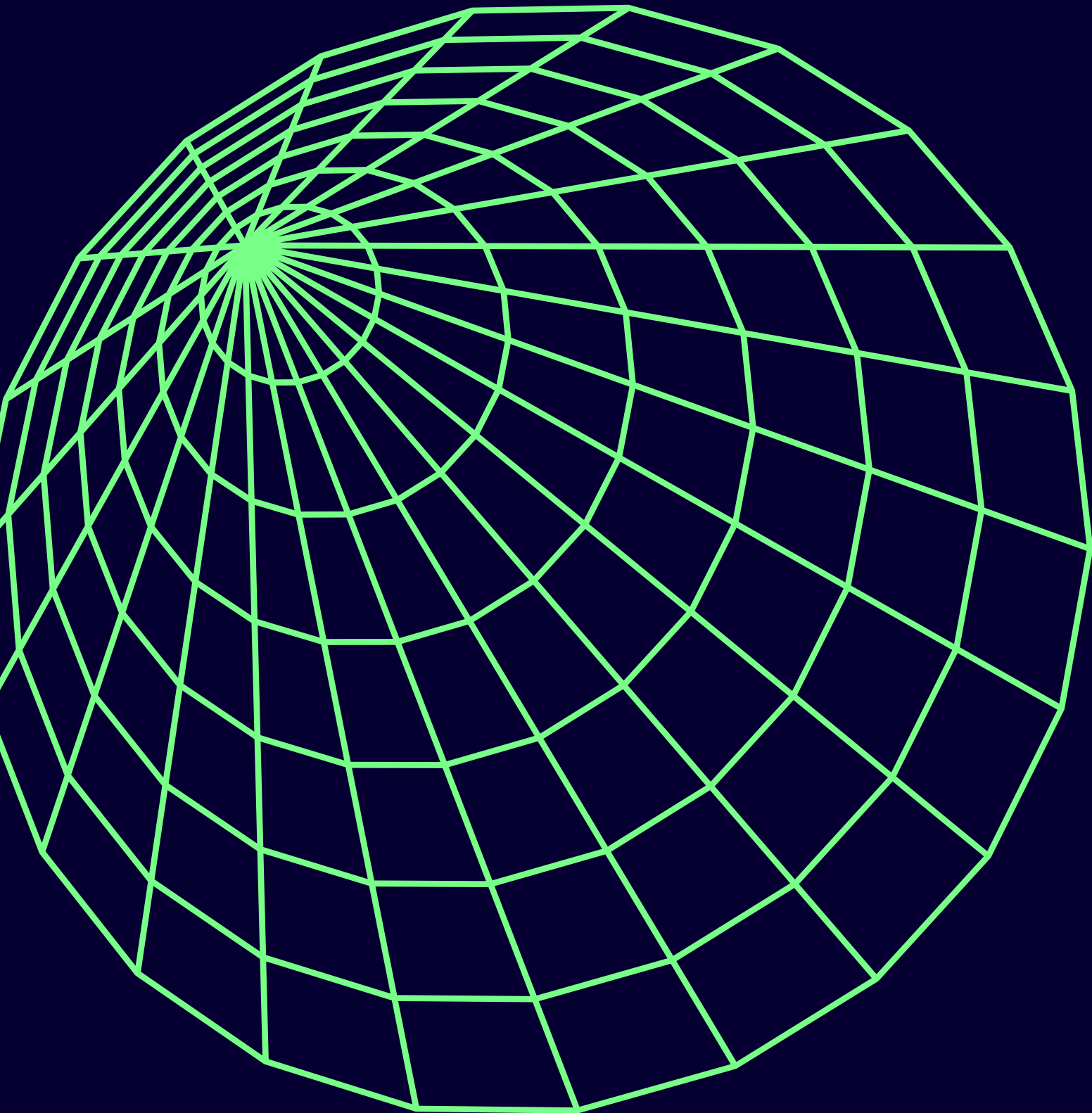
In the paper sales prediction using ML by Hitsh et al. they focused in the retail sales market and they collected Dataset from 2015-2017 which too old and also for retail channel it's the smallest it will better if they got data for wholesalers or Key Accounts (KA) the algorithm used linear regression, random forest and GBA.

Research Methodology



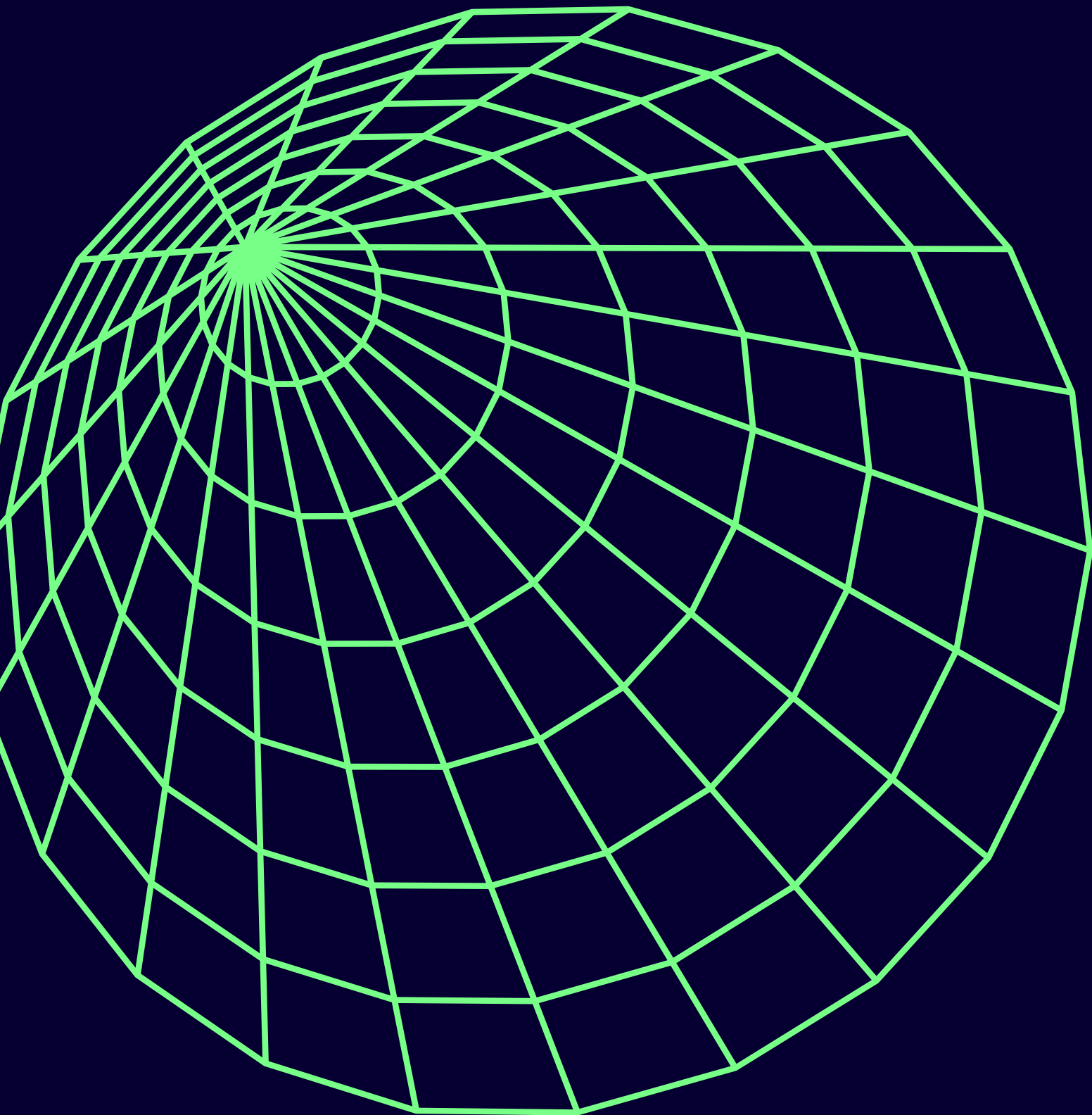
ALGORITHM	ACCURACY	RMSE
Linear Regression	0.58	11.787
Random Forest Regression	0.649	10.821
Gradient boosting algorithm	0.69	10.343

Deep learning



In the paper Retail Sales Forecasting Using Deep Learning: Systematic Literature Review (2022) Linda Eglite et al. machine learning models, including linear regression, random forest, XGBoost, and LSTM, data set 10 retail stores RMSE, MAE and R-squared to evaluate their model and XGBoost shows the best results

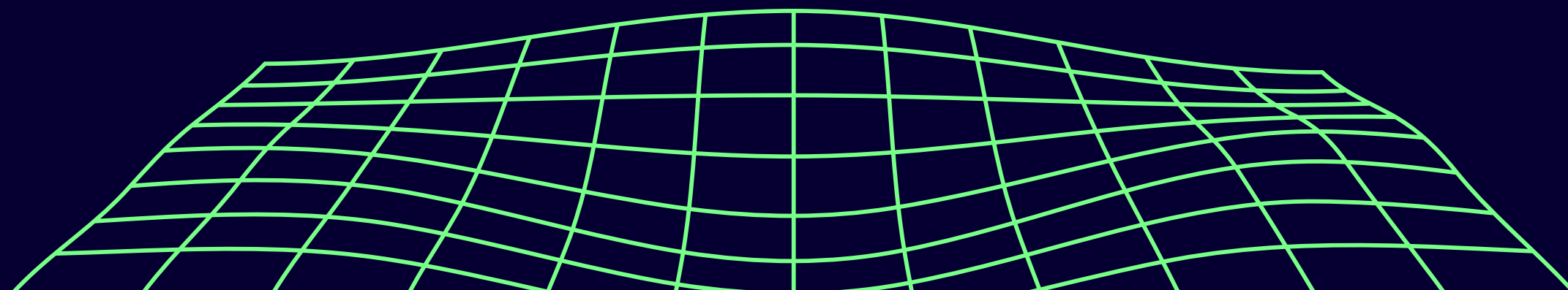
Deep Learning



In the paper Approaching Sales Forecasting Using Recurrent Neural Networks and Transformers (2022) Iván Vallés-Pérez et al. also focused on the retail sales market Dataset from 2013-2017 they provided a trick they named the “random max time step trick,” they used seq2seq and transformer architecture.

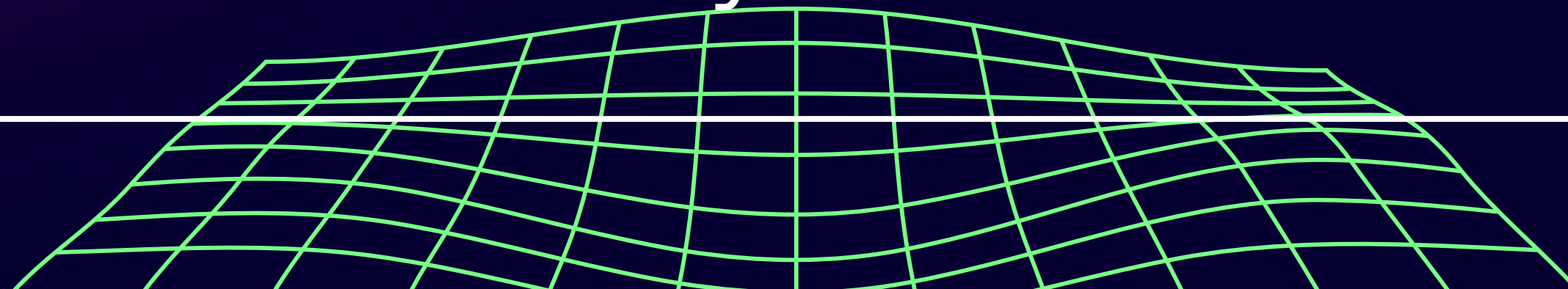
Methods

XGBoost combines multiple weak learners (typically decision trees) to create a strong predictive model. It uses gradient boosting, where each tree corrects the errors of the previous trees.



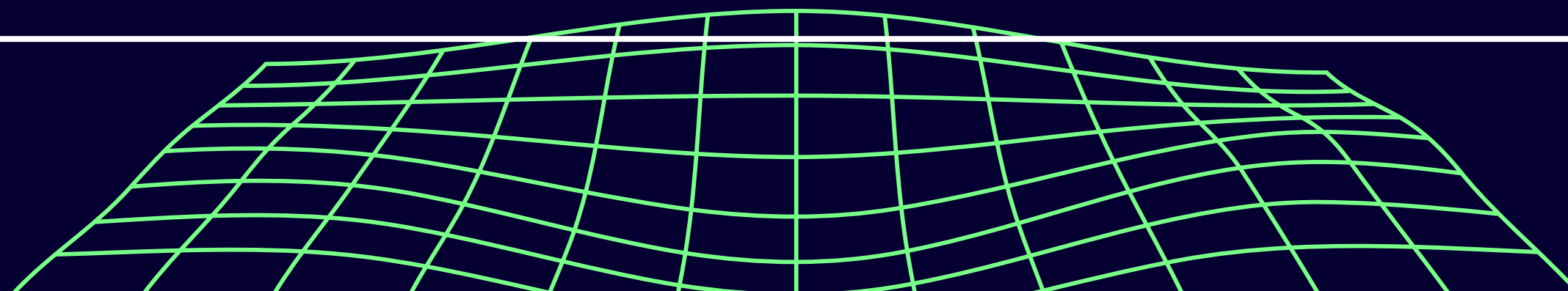
Methods

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture commonly used in machine learning, especially for sequence prediction tasks. Unlike traditional RNNs, LSTMs are designed to address the issue of vanishing gradients, allowing them to capture long-term dependencies in sequential data effectively.



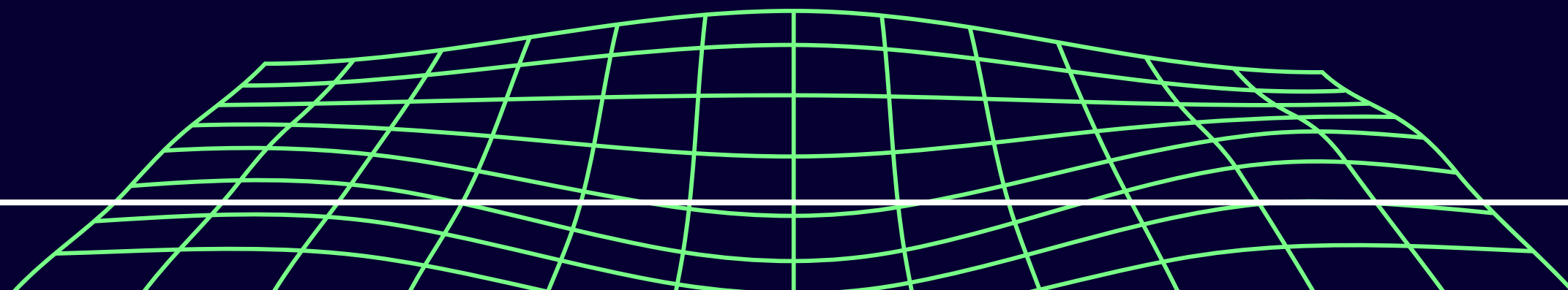
Methods

Gradient Boosting Algorithm is a machine learning technique for regression and classification problems. It builds a strong predictive model by combining multiple weaker models, typically decision trees, iteratively. Here's how it works



Methods

Random Forest is a popular machine learning algorithm primarily used for both classification and regression tasks

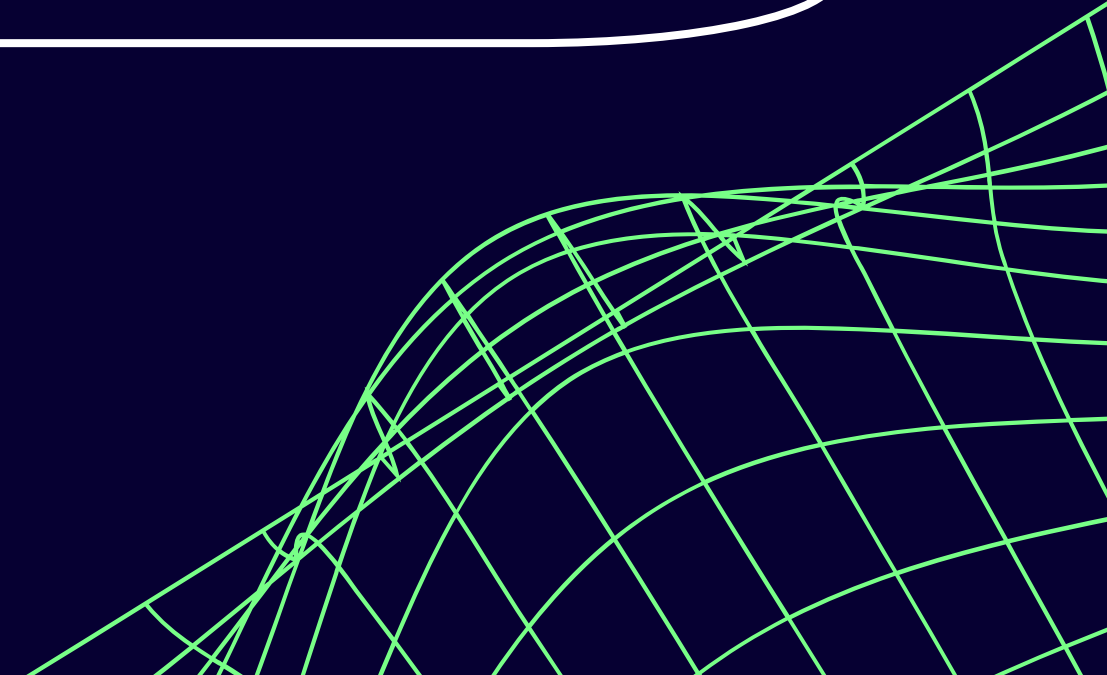


Evaluations Metrics

RMSE (Root Mean Squared Error)

Measures the square root of the average squared differences between predicted values and actual values.

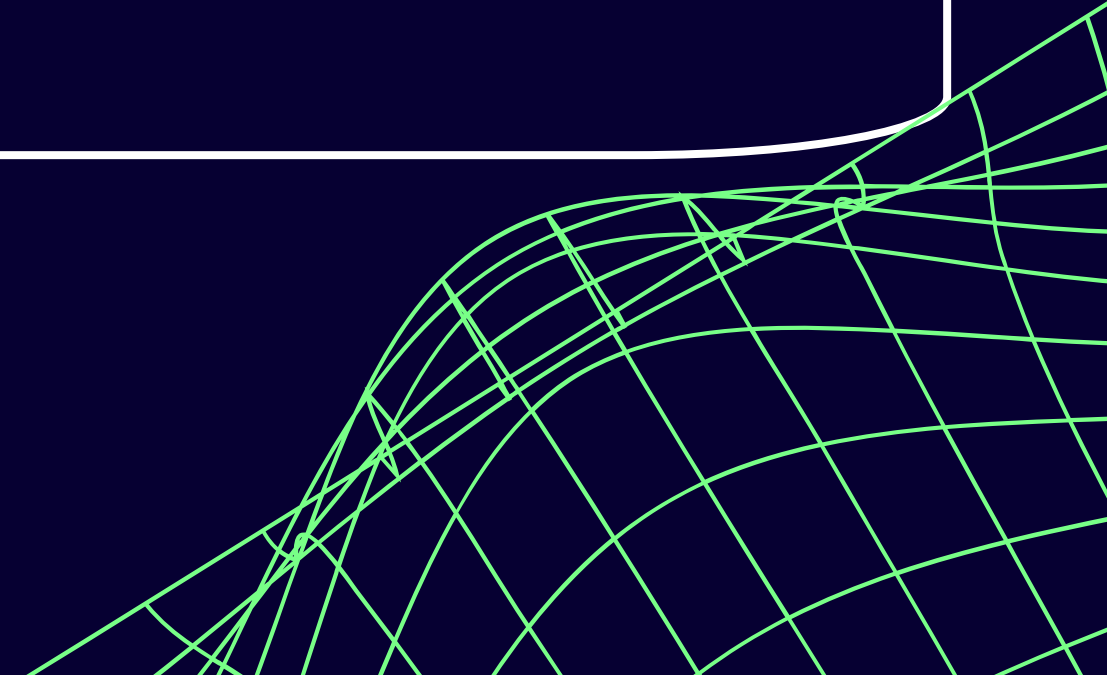
Lower RMSE is the best.



Evaluations Metrics

MAE (Mean Absolute Error)

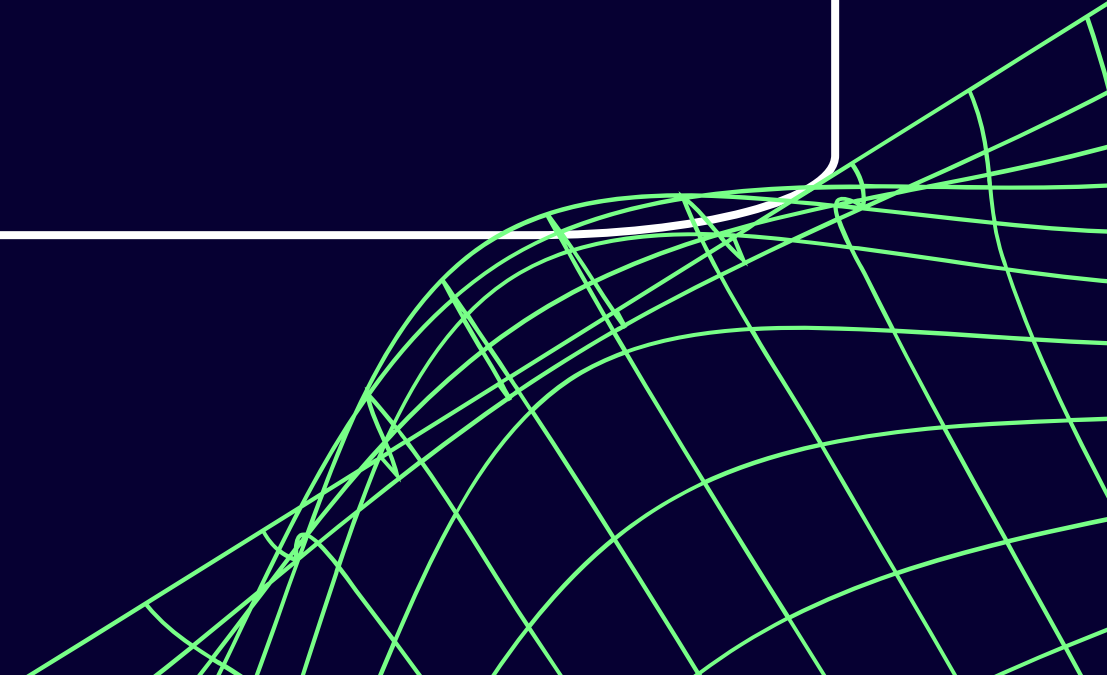
Calculates the average absolute differences between predicted values and actual values.
Lower is the best



Evaluations Metrics

MSE (Mean Squared Error)

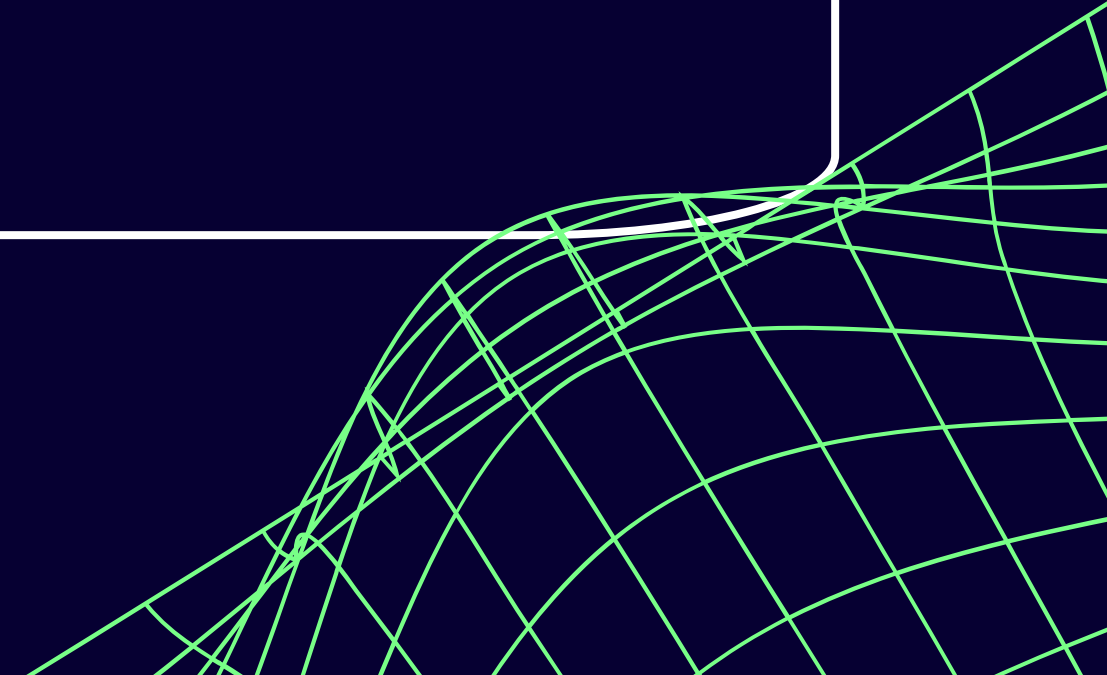
It calculates the average of the squared differences between predicted and actual values. Squaring the errors ensures all differences are positive and emphasizes larger errors.



Evaluations Metrics

MSE (Mean Squared Error)

It calculates the average of the squared differences between predicted and actual values. Squaring the errors ensures all differences are positive and emphasizes larger errors.



The End

