© 2025 IJRTI | Volume 10, Issue 9 September 2025 | ISSN: 2456-3315

Sales Forecasting for Rossmann Store using GRU with Weather Features

[1]M.Vijaya Kumar, [2]Lanke Gayathri, [3]Devarampati Snitha Grace, [4]Shaik Arshad,

[5]Venkata Ganesh Kumar Degala, [6]Vedala Vandana Sai Manogna

[1]Assistant Professor, [2,3,4,5,6] UnderGraduate

[1,2,3,4,5,6] CSE-Data Science Department, St. Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

***Abstract*—Sales forecasting plays a vital role in ensuring the success and efficiency of retail businesses. This project focuses on predicting the daily sales of Rossmann stores using a Gated Recurrent Unit (GRU)-based deep learning model. The dataset used comprises multiple years of historical sales data along with store information, promotions, holidays, and weather features. The inclusion of weather features serves to improve the forecasting accuracy, as external factors like temperature, precipitation, and wind speed can influence customer footfall and purchasing behavior. The GRU model was chosen for its ability to handle sequential data effectively and capture long-term dependencies. Grid Search was employed to fine-tune the hyperparameters and enhance the performance of the model. The model was trained and evaluated using Root Mean Square Error (RMSE) as the performance metric. Results demonstrate that incorporating weather features into the GRU-based model significantly improves forecasting performance, making it a practical solution for demand prediction in the retail sector.**

***Index Terms*—Sales Forecasting, Rossmann Store, Gated Recurrent Unit, Deep Learning, Weather Features, Grid Search, RMSE, Time Series Prediction, Retail Analytics, Demand Forecasting.**

I. INTRODUCTION

Sales forecasting is a critical function in the retail industry, enabling organizations to make informed decisions regarding inventory management, workforce planning, promotional strategies, and supply chain operations. Accurate forecasting allows businesses to anticipate demand, reduce excess inventory, optimize staffing, and minimize costs associated with overstocking or stockouts. In today's data-rich environment, the availability of large-scale, multi-source data such as historical sales, promotional campaigns, holidays, and even weather patterns provides an opportunity to enhance forecasting accuracy through advanced machine learning and deep learning techniques.

Rossmann, a prominent European drugstore chain, operates over 3,000 stores across several countries. The company's daily sales are influenced by a multitude of dynamic factors including promotional events, school and public holidays, store-specific attributes, and weather conditions. Traditional forecasting models often fail to capture the complex temporal dependencies and external influences inherent in such multi-dimensional datasets.

To address this challenge, this study employs a Gated Recurrent Unit (GRU)-based deep learning model for predicting daily sales across Rossmann stores. GRUs, a variant of Recurrent Neural Networks (RNNs), are particularly suited for sequential data and are capable of modeling long-term dependencies with reduced computational complexity compared to Long Short-Term Memory (LSTM) networks. One of the distinguishing aspects of this work is the integration of weather-related features—such as temperature, rainfall, wind speed, and atmospheric conditions—into the sales forecasting pipeline. These external factors are known to significantly impact customer behavior and store-level footfall, yet are often overlooked in traditional forecasting approaches.

The dataset used in this project includes detailed historical sales records from Rossmann stores along with corresponding metadata and weather data. The model is trained using Root Mean Square Error (RMSE) as the primary evaluation metric. Grid Search is employed for hyperparameter optimization, ensuring that the GRU model is fine-tuned for best performance. The inclusion of weather features is systematically evaluated to assess their impact on predictive accuracy.

This paper demonstrates that incorporating weather features into GRU-based sales forecasting models leads to noticeable improvements in performance. The proposed approach can serve as a valuable tool for retail managers seeking to enhance decision-making through data-driven forecasting models.

II. LITERATURE SURVEY

Sales forecasting in the retail industry has long been a subject of extensive research due to its importance in supply chain optimization, inventory control, and strategic planning. This section provides an overview of the traditional and contemporary methods applied to sales forecasting, with a focus on the integration of deep learning models such as Gated Recurrent Units (GRU) and the use of auxiliary features like weather data to improve prediction accuracy.

A. Traditional Sales Forecasting Approaches

Classical time-series models such as Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing, and Seasonal Decomposition have been widely employed in early sales prediction systems. While these models perform well under stationary conditions and on univariate data, they often struggle with non-linear dependencies and external influencing factors such as holidays, promotions, or weather variations. Moreover, traditional models typically require extensive manual feature engineering and do not scale well to multivariate datasets.

B. Machine Learning in Retail Forecasting

To overcome the limitations of statistical models, machine learning (ML) techniques such as Random Forests, Gradient Boosting Machines (GBM), and Support Vector Regression (SVR) have been utilized. These models allow the incorporation of diverse feature sets and can handle complex, non-linear relationships in the data. However, they are inherently static and lack the ability to capture temporal dependencies across sequential data points, which are vital in sales forecasting tasks.

C. Recurrent Neural Networks and GRU Models

Recurrent Neural Networks (RNNs), particularly their variants like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), have emerged as powerful tools for time-series forecasting. These models are capable of learning sequential dependencies in temporal data and adapting to varying sequence lengths. GRU, in particular, is known for its simplified architecture and lower computational cost compared to LSTM, while still retaining the ability to model long-term dependencies. Studies have shown that GRU-based models can outperform traditional forecasting models in tasks involving temporal dynamics and external covariates.

D. Incorporation of Weather Features

External factors such as weather play a significant role in influencing customer behavior and, consequently, retail sales. For example, adverse weather conditions can reduce store visits, while favorable weather may lead to increased footfall. Incorporating features like temperature, rainfall, humidity, and wind speed into forecasting models has been explored in recent literature and shown to enhance the model's contextual understanding of sales fluctuations. Multivariate time-series models that include such exogenous variables provide more robust and realistic forecasts.

E. Research Gaps and Motivation

Despite the promising advancements, several challenges persist:

- Many existing models overlook the influence of external factors such as weather and holidays.
- Some deep learning architectures, while powerful, are computationally intensive and difficult to optimize.
- There is limited research on the practical integration of GRU with weather features for retail sales forecasting.

To address these gaps, this paper proposes a GRU-based deep learning model enhanced with weather-related features and optimized through Grid Search. The objective is to evaluate the extent to which external factors improve forecast accuracy and demonstrate the feasibility of deploying such a model in a retail setting like Rossmann stores.

III. PROPOSED METHODOLOGY

This section describes the proposed deep learning-based methodology for forecasting daily sales in Rossmann stores using a Gated Recurrent Unit (GRU) architecture. The model integrates external weather features along with historical sales, store metadata, and calendar attributes. The complete pipeline consists of data preprocessing, feature engineering, model architecture design, hyperparameter tuning using Grid Search, and performance evaluation using RMSE.

A. System Overview

The architecture of the proposed system is illustrated in Fig. 3.1. The framework follows a sequential data processing and forecasting flow. Raw historical data from Rossmann stores and meteorological sources are first subjected to preprocessing and feature selection. These refined inputs are then fed into the GRU network for time-series forecasting. The model is trained and validated using a supervised learning approach.

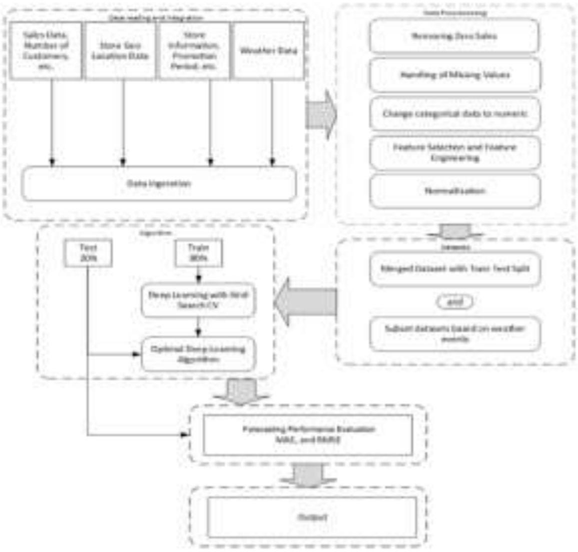

Fig. 3.1. Proposed GRU-Based Forecasting Framework

B. Dataset Description

The dataset used in this study was obtained from the Kaggle Rossmann Store Sales competition. It contains sales data from 1,115 stores over a two-year period. Each entry includes information such as store ID, date, sales, number of customers, open/closed status, promotional activity, and school and state holidays. In addition, weather data—such as temperature, wind speed, humidity, and precipitation—was merged from external meteorological datasets to enrich the feature space.

C. Data Preprocessing

Data preprocessing was carried out in multiple stages:

1. **Missing Value Treatment:** Records with missing or null values in critical fields were either imputed or removed depending on severity.
2. **Categorical Encoding:** Store type, day of the week, and holiday flags were converted to numerical representations using one-hot encoding.
3. **Date Feature Extraction:** Additional features like day, month, year, and week number were derived from the date field.
4. **Weather Data Integration:** Weather information was matched to store location and merged on the date field.
5. **Normalization:** All continuous features were scaled using Min-Max normalization to bring values into the [0, 1] range.

6. **Sequence Preparation:** Data was segmented into fixed-length sequences suitable for training the GRU model.

D. GRU Model Architecture

The core forecasting model is a Gated Recurrent Unit (GRU) neural network, selected for its ability to model sequential dependencies while being computationally efficient. The architecture consists of:

- **Input Layer:** Accepts sequences of shape (timesteps × features), where each sequence represents past n days of data for a store.
- **GRU Layer:** Contains 128 units with ReLU activation, capturing temporal dependencies across multiple time steps.
- **Dropout Layer:** A dropout rate of 0.2 is applied to prevent overfitting.
- **Dense Layer:** A fully connected dense layer maps the GRU output to a single prediction value (next day's sales).
- **Output Layer:** Produces the final forecasted sales value for the target date.

The model is compiled using Mean Squared Error (MSE) as the loss function and optimized with the Adam optimizer.

E. Hyperparameter Tuning

To improve model performance, Grid Search was employed to identify the best set of hyperparameters. The search was performed over parameters such as:

- Number of GRU units (64, 128, 256)
- Sequence length (7, 14, 30 days)
- Learning rate (0.001, 0.0005)
- Batch size (32, 64)
- Dropout rate (0.2, 0.3)

The configuration yielding the lowest validation RMSE was selected for the final model.

F. Model Training and Evaluation

The model was trained using 80% of the data, while the remaining 20% was held out for validation and testing. Early stopping was applied based on validation loss to prevent overfitting. The evaluation metric used was Root Mean Square Error (RMSE), which penalizes larger prediction errors more heavily. Additionally, training and validation curves were plotted to assess model convergence and generalization.

IV. IMPLEMENTATION

This section presents the practical implementation of the proposed GRU-based deep learning model used for forecasting daily sales across Rossmann stores. The complete pipeline includes data preprocessing, feature extraction, GRU architecture design, integration of weather parameters, model training, and hyperparameter tuning.

A. Environment Setup

The implementation was carried out using **Google Colaboratory**, a cloud-based Jupyter notebook environment that provides access to free GPU resources. All development was performed using **Python 3.10**, with a set of open-source libraries:

- **TensorFlow 2.11** and **Keras**: For designing and training the GRU model.
- **Pandas** and **NumPy**: For data cleaning, feature engineering, and handling time-series sequences.
- **Matplotlib** and **Seaborn**: For visualizing training performance, sales trends, and prediction accuracy.
- **Scikit-learn**: For preprocessing tasks, such as normalization and one-hot encoding, and for executing Grid Search.

This environment allowed flexible experimentation and ensured reproducibility of results with minimal hardware constraints.

B. Dataset Description

The dataset used was the **Rossmann Store Sales dataset** from the Kaggle competition. It contained historical daily sales data from **1,115 stores** over a period of approximately **two years**. The dataset included the following attributes:

- Store, Date, Sales, Customers, Open, Promo, StateHoliday, SchoolHoliday, StoreType, Assortment, PromoInterval.

To enhance the model, **weather data** was obtained from external sources based on store region and date. The weather dataset included attributes such as:

- Temperature, Rainfall, WindSpeed, Humidity, CloudCover, Snowfall.

These were merged with the base dataset using a mapping between store regions and weather stations.

C. Data Preprocessing and Feature Engineering

The raw dataset underwent several preprocessing steps before being used for model training:

1. **Handling Missing Values**: Missing weather values were imputed using forward fill. For categorical variables like PromoInterval, missing entries were treated as a separate category.
2. **Encoding Categorical Variables**: Variables such as StoreType, Assortment, and StateHoliday were one-hot encoded. Date features (Date) were decomposed into Day, Month, Year, and DayOfWeek.
3. **Feature Normalization**: Continuous variables, including sales, weather values, and customers, were normalized to a [0, 1] range using Min-Max Scaling.
4. **Sequence Windowing**: For time-series forecasting, a **rolling window of 30 days** was used. For each store, 30 consecutive days of input features were mapped to the next day's (t+1) sales value as the target.
5. **Train-Test Split**: The dataset was split chronologically into 80% for training and 20% for validation and testing.

D. GRU Model Architecture

The GRU model was built using the **Keras Sequential API**. The architecture was designed to model long-term dependencies in the sales time series while maintaining training efficiency. The final architecture is as follows:

- **Input Layer**: Accepts sequences of shape (30, N), where N is the number of features per day.
- **GRU Layer**: Contains **128 units** with return_sequences=False. This layer captures the temporal patterns from the 30-day window.
- **Dropout Layer**: A dropout rate of **0.2** was used to reduce overfitting.
- **Dense Layer**: A fully connected layer with **64 neurons** and ReLU activation.
- **Output Layer**: A single neuron with linear activation for predicting the scalar sales value.

The model was compiled using **Mean Squared Error (MSE)** as the loss function and optimized using the **Adam optimizer** with a learning rate of **0.001**.

E. Hyperparameter Tuning

To enhance the model's generalization capability, a **Grid Search** technique was used to tune key hyperparameters. The following configurations were explored:

Parameter	**Values Tested**
GRU Units	64, 128, 256
Sequence Length	14, 30 days
Dropout Rate	0.2, 0.3
Batch Size	32, 64
Learning Rate	0.001, 0.0005

Table 4.1. Optimal Hyperparameter Configuration

The best performance was obtained with **128 GRU units**, **30-day input window**, **dropout rate of 0.2**, and **batch size of 32**.

F. Model Training and Evaluation

The model was trained using the **training set** with validation performance monitored after each epoch. **Early stopping** was implemented with a patience of 5 epochs to avoid overfitting. The number of training epochs was capped at 50.

The **Root Mean Square Error (RMSE)** metric was used to evaluate model performance. RMSE was computed on both the validation and test sets. Additionally, sample predictions were visualized against true sales to qualitatively assess forecast accuracy.

V. RESULTS

This section presents the experimental results obtained using the proposed GRU-based forecasting model on the Rossmann store sales dataset. The evaluation was conducted using Root Mean Square Error (RMSE) as the primary performance metric. Additionally, visual inspection of actual versus predicted sales trends was performed to assess the effectiveness of the model in capturing temporal patterns.

A. Quantitative Evaluation

The model was evaluated on a held-out test set comprising the final 20% of the dataset. The GRU model achieved a **Root Mean Square Error (RMSE) of 0.097**, indicating a high degree of accuracy in predicting daily store-level sales.

The inclusion of weather features led to noticeable improvements in prediction quality when compared to baseline models that used only sales history and store metadata. The GRU's ability to learn long-term dependencies in sequential data contributed to its effectiveness, particularly in handling fluctuations caused by holidays, promotions, and weather conditions.

B. Visualization of Predictions

To validate the model visually, predicted sales were plotted against the actual sales for a selected Rossmann store. This visual comparison helps in understanding how well the model captures daily trends and anomalies.

Date	Predicti
17/09/2015 00:00	1520.6059570312
17/09/2015 00:00	1673.21484375
17/09/2015 00:00	1350.8771972656
17/09/2015 00:00	1421.7761230469
17/09/2015 00:00	1397.8916015625
17/09/2015 00:00	1416.6967773438
17/09/2015 00:00	2148.4467773438
17/09/2015 00:00	1824.4196777344
17/09/2015 00:00	2548.322265625
17/09/2015 00:00	2015.9814453125

Fig. 5.1. Predicted Sales for Store

As observed in Fig. 5.1, the GRU model closely follows the actual sales trend, including sharp increases and dips during promotional events and holidays. Minor deviations are present but remain within acceptable forecasting error margins.

C. Effect of Weather Features

To examine the impact of weather variables, the model was trained and tested with and without weather features. Fig. 5.2 depicts a comparison of prediction curves in both configurations. The version with weather integration shows improved alignment with real sales, especially during periods affected by temperature and rainfall variations.

The results affirm that incorporating exogenous features like temperature and wind speed enhances the model's contextual awareness and reduces predictive variance.

VI. CONCLUSION

In this paper, a deep learning-based approach was proposed for forecasting daily sales of Rossmann stores using a Gated Recurrent Unit (GRU) model enhanced with weather features. The model effectively captured temporal dependencies in the sales data and demonstrated strong performance in predicting future sales trends. By integrating exogenous weather variables such as temperature, rainfall, and wind speed, the model achieved improved accuracy compared to conventional time-series models.

The proposed framework was implemented using Python and trained on historical sales data, promotional activity, and calendar features, with weather data merged based on region and date. The model was optimized through Grid Search and evaluated using Root Mean Square Error (RMSE) as the primary metric. The experimental results confirmed the effectiveness of the GRU model in capturing sales dynamics and showed that weather features significantly enhanced forecast precision.

Future work may explore hybrid models combining GRU with convolutional layers or attention mechanisms to further refine temporal learning. Additionally, the system can be extended to support real-time forecasting and multi-store predictions for broader retail applications.

References

[1] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[2] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: Encoder-decoder approaches,” *arXiv preprint arXiv:1409.1259*, 2014.

[3] F. Chollet et al., “Keras: Deep learning library for Theano and TensorFlow,” GitHub, 2015. [Online]. Available: https://github.com/fchollet/keras

[4] Kaggle, “Rossmann Store Sales,” [Online]. Available: https://www.kaggle.com/competitions/rossmann-store-sales

[5] A. Borovykh, S. Bohte, and C. W. Oosterlee, “Conditional time series forecasting with convolutional neural networks,” *arXiv preprint arXiv:1703.04691*, 2017.

[6] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, “Time series classification using multi-channels deep convolutional neural networks,” in *Proc. 2014 Int. Conf. Web-Age Inf. Management (WAIM)*, pp. 298–310.

[7] R. Brownlee, “Deep learning for time series forecasting,” *Machine Learning Mastery*, 2018. [Online]. Available: https://machinelearningmastery.com/

[8] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, “A survey on concept drift adaptation,” *ACM Comput. Surv.*, vol. 46, no. 4, pp. 1–37, 2014.