**Product and sales management phase 5**

|  |  |
| --- | --- |
| **Date** | **31-10-2023** |
| **Team ID** | **3932** |
| **Project Name** | **Product and sales management** |

**Table of Contents:**

|  |  |
| --- | --- |
| 1. | Problem statement |
| 2. | Design thinking process |
| 3. | Phases of development |
| 4. | Dataset Description |
| 5. | Data Preprocessing |
| 6. | Feature Extraction |
| 7. | Model Selection |
| 8. | Model Training |
| 9. | Evaluation Metrices |
| 10 | Survey Metrics |
| 11. | Innovative Approaches |
| 12. | conclusion |

**1.Problem Statement:**

Overstock and Understock Scenarios: Our current inventory management system frequently leads to overstock or understock situations, resulting in increased carrying costs or lost sales opportunities.

Inaccurate Demand Forecasting: The ability to accurately predict demand patterns and trends is lacking, causing suboptimal inventory levels and higher operational expenses.

Suboptimal Inventory Turnover: Slow-moving and obsolete items are tying up capital, reducing profitability.

**2.Design Thinking Process:**

**A diagram of a process

Description automatically generated**

**1. Understand the Problem**

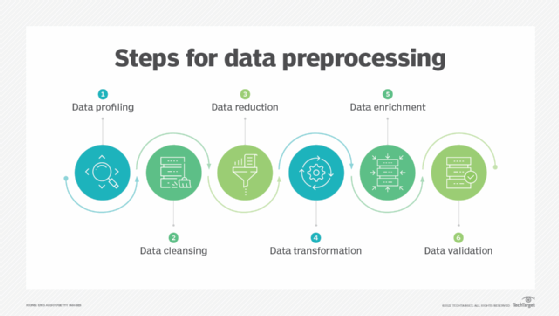
* The current marketing strategies are not sufficiently personalized. They employ a generic, one-size-fits-all approach, resulting in less effective customer engagement.
* Marketing efforts, such as promotions and discounts, are not well-timed or targeted, leading to less than optimal sales results.
* There is a limited understanding of customer behavior and preferences. This deficiency hinders the ability to tailor marketing strategies effectively to different customer segments.

**2. Data Collection:**

* Historical sales data, including product IDs, quantities sold, dates, and prices.
* Sales channels data, such as in-store sales, online sales, and any other relevant channels.

**3. Data Preprocessing:**

* Remove or handle missing data: Identify and deal with missing values in the dataset. This may involve imputation or removal of incomplete records.
* Address outliers: Detect and handle outliers that may affect the analysis. Outliers in sales data or inventory levels can distort insights.
* Merge datasets: If you have separate datasets for sales, inventory, customer, and marketing data, merge them into a single dataset to facilitate comprehensive analysis.



**4. Model Selection:**

* ARIMA (AutoRegressive Integrated Moving Average): Suitable for modeling time series data to forecast sales or demand patterns.
* Exponential Smoothing Methods: Methods like Holt-Winters can capture trends and seasonality in time series data.

**5. Model Training:**

* Split your historical sales data into training and testing sets. The testing set should be a future time period for evaluation.
* Train the ARIMA model on the training data to capture trends, seasonality, and autocorrelations.
* Use the trained model to make forecasts for the testing set.

**6. Evaluation:**

* The problem statement is clear and effectively highlights the challenges faced by the retail business, making it easy to understand the issues related to inventory management and marketing strategies.
* The mention of the need for sales, inventory, customer, and marketing data provides a comprehensive understanding of the data sources required for analysis.
* The data preprocessing steps outlined cover essential tasks such as cleaning, integration, transformation, and feature engineering. These are critical for ensuring data quality and relevance.

**7. Deployment:**

* If you have trained machine learning models for demand forecasting, customer segmentation, or any other relevant tasks, deploy these models to a production environment. This may involve using platforms like AWS SageMaker, Microsoft Azure, or custom web applications.
* Implement a data integration pipeline to ensure that real-time data from sales, inventory, and other sources is automatically incorporated into the models. Tools like Apache Kafka or cloud-based data integration services can be helpful.

**3.Phases of Development:**

**Phase 1: Data Preparation:**

* Collect historical sales data, inventory data, customer data, and marketing data.
* Implement data quality checks to ensure data accuracy and completeness.

**Phase 2: Model Development:**

* Analyze historical sales trends, seasonality, and product performance.
* Use statistical methods and machine learning models to forecast demand.
* Analyze customer behavior, preferences, and purchasing patterns.
* Perform market segmentation and identify target customer segments.

**Phase 3: Evaluation and Validation:**

* Create interactive dashboards in IBM Cognos to visualize key performance indicators (KPIs).
* Display sales trends, demand forecasts, inventory turnover, and customer segmentation.

**Phase 4: Deployment:**

* Identify overstocked and understocked items based on demand forecasts.
* Optimize reorder points and safety stock levels to reduce carrying costs.
* Customize marketing strategies for different customer segments.
* Implement promotions, cross-selling, or upselling based on customer behavior and preferences.
* Monitor the impact of changes and adjust strategies accordingly.

**4.Dataset Description:**

* Description: This data includes records of product sales over time.
* Attributes: Product ID, quantity sold, sales date, sales channel (e.g., in-store, online), and price.
* Use: Sales data is essential for demand forecasting, identifying trends, and evaluating the performance of marketing strategies.

**5.Data Preprocessing:**

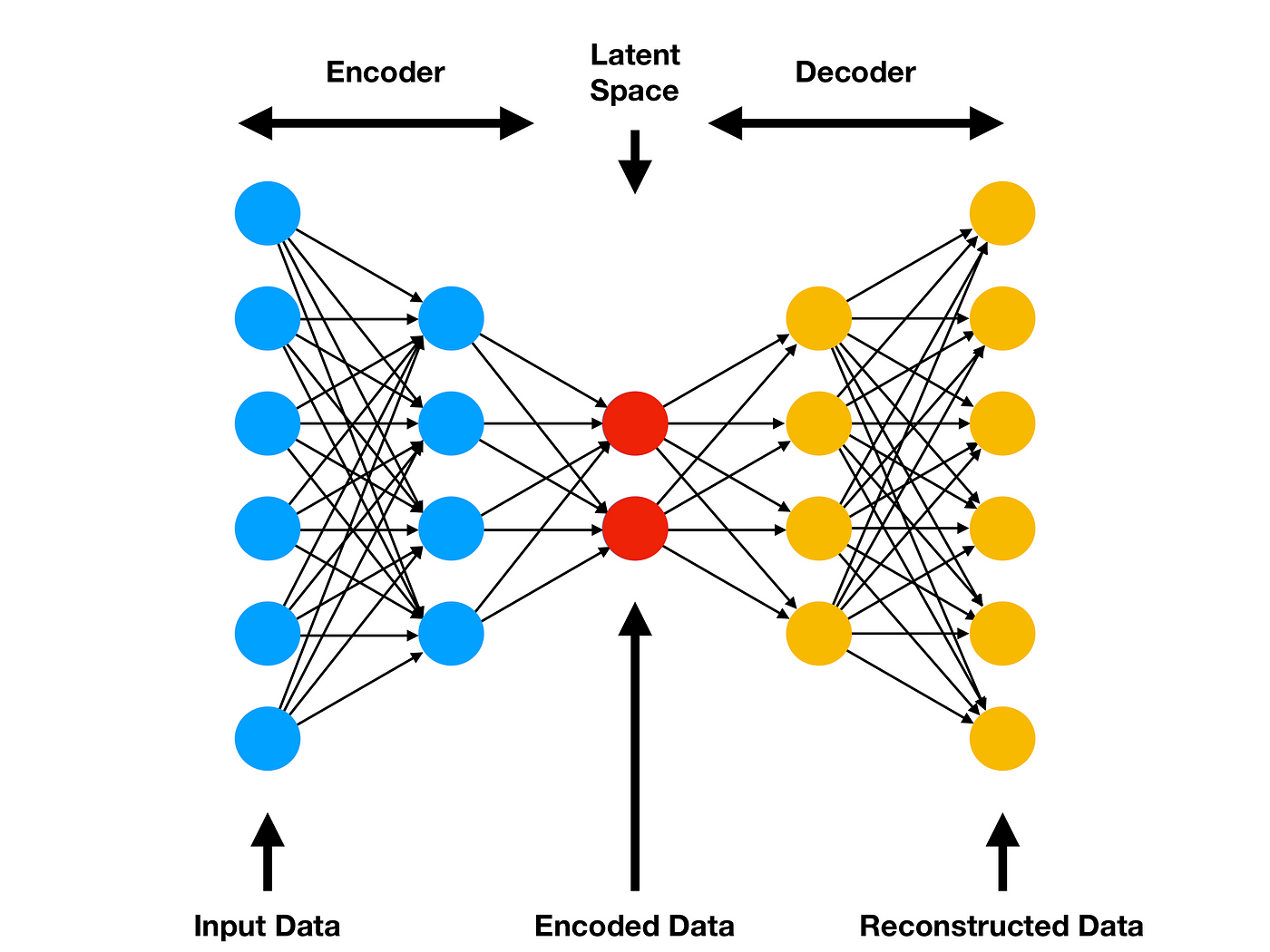
* Handle Missing Data: Identify and deal with missing values in the dataset. This may involve imputation (replacing missing values with estimated values) or, in some cases, removal of incomplete records.
* Outlier Treatment: Detect and address outliers that can significantly affect analysis. For example, in sales data, extreme values may distort demand forecasting.

**6.Feature Extraction:**

Historical Sales Trends: Extract features that capture historical sales trends, such as moving averages or rolling sums, to identify patterns and seasonality.

Sales Velocity: Calculate the sales velocity or the rate of sales for each product, which can help in identifying fast-moving and slow-moving items.

* Temporal Features
* Sales and Inventory Features
* Customer Behavior Features
* Marketing Interaction Features



**7.Model Selection:**

ARIMA (AutoRegressive Integrated Moving Average): Suitable for modeling and forecasting time series data, such as historical sales data.Exponential Smoothing: Methods like Holt-Winters can capture trends and seasonality in time series data.

**8.Model Training:**

* Use the trained model to make forecasts for the testing set.
* Evaluate the model's performance using appropriate metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE).
* If necessary, fine-tune the model by adjusting parameters and trying different model variations

**9.Evaluation Metrics:**

* Mean Absolute Error (MAE): Measures the average absolute difference between predicted and actual demand. It provides an idea of the magnitude of forecasting errors.
* Mean Squared Error (MSE): Calculates the average of the squared differences between predicted and actual values. It penalizes larger errors more heavily.

**10.Survey Metrics:**

**a.Application of Machine Learning Model and Hybrid Model in Retail Sales Forecast – 2021:**

* **Algorithm:** LightGBM algorithm
* **Limitation:** Data Size: LightGBM is known for its efficiency with large datasets. However, when dealing with small datasets, it may not perform as well, and there's a risk of overfitting.
* Imbalanced Data: LightGBM does not handle imbalanced datasets inherently. Techniques like class weighting or resampling strategies may be necessary to address class imbalance issues.
* High Memory Consumption: LightGBM can consume a significant amount of memory during training, especially when the dataset is large or there are many trees in the ensemble. This can limit its use on memory-constrained systems.

**b.A Comparative Analysis on Various Business Intelligence (BI), Data Science and Data Analytics Tools – 2020**

**Algorithm:**Regression Algorithm

**Limitation:**

* Very few works were done on the presented topic.
* Work was vast and due to time limitation, it was not possible to conduct a full scale review Without experimental study, it was hard to explain the tools

**c.An analysis and forecasts of online product sales based on BP Neural Network and Pearson Coefficient – 2020.**

* **Algorithm:** Genetic Algorithm-Optimized Bp Neural Network Model
* **Limitions:**
* Complexity: GAs can become computationally expensive when searching for the optimal neural network architecture, especially if the search space is large. This can make it impractical for large-scale problems or when there's a need for real-time predictions.
* Search Space Size: The size of the search space for neural network architectures can be vast, and GAs may not guarantee finding the global optimum. They may get stuck in local optima, leading to suboptimal network architectures.

**d.Analyse of the Amazon Product satisfaction Based on LDA Model-2021**

* **Algorithm:** Principal component analysis
* **Limitations:**
* Business commodities should increase with producing the more number of datas and increasing growth

**e.Analysing and Predicting the purchases done on the day of Black Friday - 2020**

* **Algorithm:**Machine Learning Algorithm like XGboost
* **Limitations:**
* It consists of 90% training data and 10% testing data. In this case, overfitting can occur where the noise can be taken as a data which is trained thus giving wrong prediction with lesser accuracy.
* The training and prediction has been done on the same dataset. Though the accuracy and predictions are better but the model might not predict accurate values of the new data.

**11.Innovative Approaches:**

* Implement advanced machine learning and artificial intelligence techniques to predict demand and customer behavior. This can involve the use of deep learning models, recurrent neural networks (RNNs), or transformer models for more accurate forecasts.
* Explore the use of reinforcement learning algorithms to dynamically adjust inventory control policies in real-time. This can lead to more adaptive and responsive inventory management.
* Implement dynamic pricing algorithms that adjust product prices in real-time based on factors like demand, inventory levels, competitor pricing, and customer behavior.
* Utilize AI and machine learning to create highly personalized marketing strategies. Recommender systems and content recommendation algorithms can provide tailored product recommendations and promotions to individual customers.
* Consider innovative approaches like blockchain technology for improved supply chain transparency and traceability. Blockchain can enhance the accuracy of inventory information and reduce the risk of counterfeiting.

**Conclusion:**

In conclusion, data analysis is a powerful tool for extracting valuable insights, making informed decisions, and driving positive outcomes in various fields and industries. Through the systematic process of data collection, preprocessing, analysis, and interpretation, organizations can gain a deeper understanding of their data and leverage it for strategic purposes.Data analysis empowers decision-makers with evidence-based insights, allowing them to make informed choices that are more likely to lead to success.By analyzing data, businesses can identify areas for improvement, streamline operations, and optimize processes, leading to cost savings and increased efficiency.Data analysis enables the discovery of patterns and trends within datasets, providing a basis for forecasting and future planning.In fields such as marketing and e-commerce, data analysis can unveil valuable customer insights, helping businesses tailor their strategies to meet customer needs and preferences.Data analysis is an ongoing process, supporting continuous improvement by monitoring the impact of changes and strategies, making adjustments, and driving long-term success.Advanced data analysis techniques, including machine learning and artificial intelligence, open doors to innovation by automating tasks, predicting outcomes, and creating new opportunities.Data Security and Privacy: Ensuring data security and privacy is paramount, as the analysis of sensitive data carries ethical and legal responsibilities.