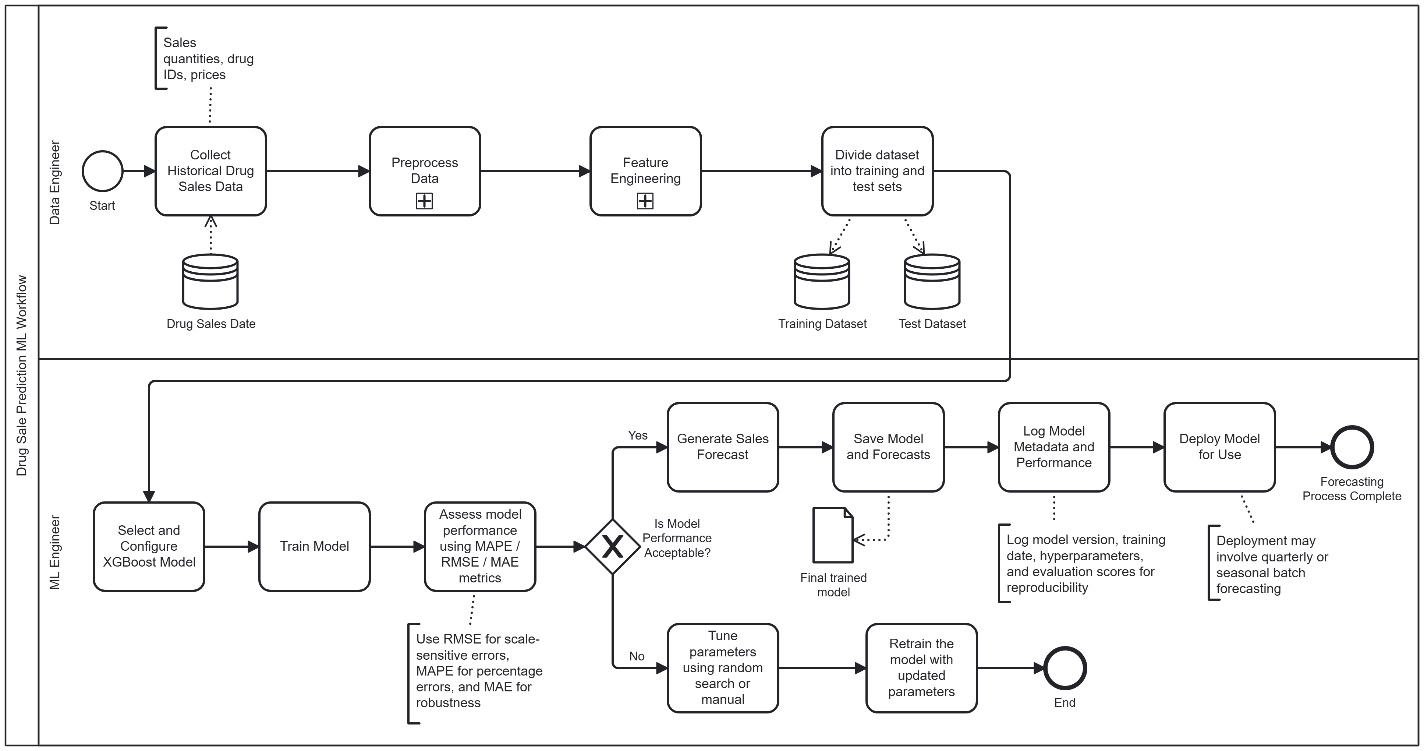
**Research on Application of machine learning methods for optimization of business processes**



**Figure 1. BPMN of the Method**

# Examples of X and Y

## **Xs of the thesis:**

* **Sales quantities:** Historical sales data (e.g., the number of units sold in previous months).
* **Drug IDs:** Unique identifiers for drugs to differentiate them.
* **Prices:** Prices at which the drugs were sold.
* **Seasonal Data:** This could include labels for each season (e.g., Winter, Spring, etc.) that help predict seasonal fluctuations in sales.
* **Month/Year:** Temporal features to help predict trends over time

## **Ys of the thesis:**

* **Next Season Sales:** The quantity of drugs expected to be sold in the next season (Winter, Spring, Summer, or Autumn). Seasonality plays a big role in drug sales, as demand for certain drugs can fluctuate with the seasons (e.g., cold medicine in winter, allergy medication in spring).
* **Monthly/Quarterly Sales Predictions:** Itrepresents the quantity of sales for each month or quarter (depending on the granularity)
* **Understanding the factors that influence sales:** depends on a combination of various factors, including:
* Historical Sales Data
* Price Changes
* Seasonality
* Promotions and Discounts
* Market Trends and External Factors

# Overview of the project

The diagram represents a workflow for predicting drug sales using Machine Learning (ML). It involves **Data Engineers** and **ML Engineers** working together to train and deploy a forecasting model (XGBoost):

**1. Data Engineer Tasks:**

1. **Collect Data**: Gather historical drug sales data (quantities, prices, drug IDs).
2. **Preprocess Data**: Clean the data (handle missing values, format it for training).
3. **Feature Engineering**: Create features like lag features (past sales) and seasonal indicators.
4. **Split Data**: Divide data into training and test sets for model training and evaluation.

**2. ML Engineer Tasks:**

1. **Select & Configure Model**: Choose and set up the XGBoost model.
2. **Train Model**: Train the model on the training dataset.
3. **Assess Performance**: Evaluate the model’s performance using metrics like MAPE, RMSE, and MAE.
4. **Generate Sales Forecasts**: After training, use the model to predict future drug sales.
5. **Save Model & Forecasts**: Store the trained model and the forecasts for future use.
6. **Log Model Data**: Keep track of model details for reproducibility (version, hyperparameters, performance).
7. **Deploy Model**: Deploy the model for ongoing predictions (e.g., quarterly, seasonal forecasting).

**End of Process:**

* **Forecasting Complete**: The model is now deployed, and the forecasting process can be repeated with new data as needed.

# Spearman Rank Corellation

A table with numbers and a number

AI-generated content may be incorrect.

**Figure 2. Spearman rank-order correlation matrix**

**Drug ID and Sale Quantity**: The correlation is weakly positive, indicating that the relationship between drug ID and sales quantity is minimal, as different drugs tend to sell in different quantities.

**Sale Quantity and Month of Sale**: There’s a slight positive correlation, suggesting that sales marginally increase throughout the month. Seasonality still influences sales, especially for seasonal products.

**Drug ID and Season of the Year**: The correlation is weak, showing that drug IDs are not strongly tied to the season, as the demand for drugs does not depend on the season in this dataset.

**Drug ID and Price of Drugs**: A weak positive correlation suggests minimal influence of drug ID on the price of drugs, indicating little variation in prices for different drugs.

**Sale Quantity and Season of the Year**: There’s a low but positive correlation, indicating a slight increase in sales in autumn compared to other seasons. Seasonality still affects sales, especially for seasonal drugs.

**Sale Quantity and Price of Drugs**: The negative correlation shows that as the price of drugs increases, the quantity sold tends to decrease, possibly due to pricing policies or demand for higher-priced drugs.

# Seasonal Trends of Drug Sales

**A graph with numbers and lines

AI-generated content may be incorrect.**

**Figure 3. Statistics on the number of sales of "Dantocept Ferurone" in the Autumn**

The chart shows that the sales of "Dantocept Ferurone" are highest in the **Autumn**, especially in 2019 and 2020. Autumn consistently sees the most sales compared to **Winter** and **Spring**, when sales are much lower. This suggests higher demand in Autumn, likely due to seasonal factors like flu season. Therefore, Autumn is the most important time for sales and stocking this medicine in pharmacies.