

Publishing Industry Analytics & Prediction System

NEURAL NETWORKS FOR THE PUBLISHING INDUSTRY:
ENHANCING CUSTOMER EXPERIENCE AND SALES

SAI VAISHNAV J

Project Title

Neural Networks for the Publishing Industry: Enhancing Customer Experience and Sales

Objective

Leverage neural networks to analyze and predict customer behavior, recommend books, and forecast demand in the publishing industry, improving both customer experience and sales.

Objectives & Goals

- Predict customer churn to improve retention strategies.
- Forecast book demand to optimize inventory management.
- Provide personalized book recommendations based on order history.
- Develop an interactive Streamlit dashboard for visualization.
- Deploy models on AWS EC2 for real-time predictions.

Project Overview

The Publishing Industry Analytics & Prediction System aims to provide insights into customer behavior, book demand, and personalized recommendations using machine learning models. This project integrates multiple data sources, performs analysis, and builds predictive models for customer churn and book demand forecasting.

Data Collection & Preprocessing

Datasets Used

- Book Data (book title, author, ISBN, publisher, etc.)
- Customer Data (customer demographics and purchase history)
- Order Data (order details, order count, revenue, etc.)

SQL Queries

- Extracting data from database to dataframe

```
query = """select * from book b
inner join publisher p on b.publisher_id=p.publisher_id
inner join book_author ba on b.book_id=ba.book_id
inner join author a on a.author_id=ba.author_id
inner join book_language bl on b.language_id=bl.language_id;"""
```

Data Cleaning & Transformation

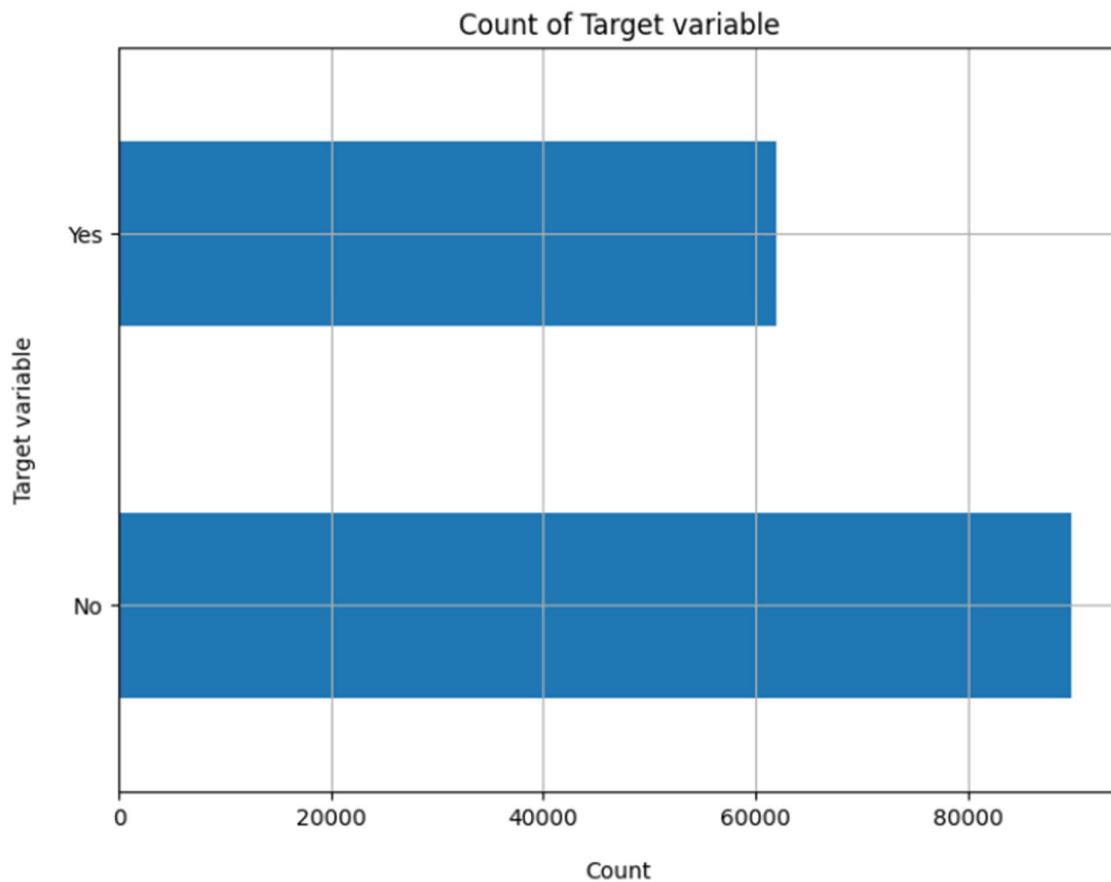
- Handling Missing Values: Replaced NaNs with suitable values.
- Feature Engineering: Created lag features & moving averages for forecasting.
- Normalization: Applied MinMaxScaler for model training.

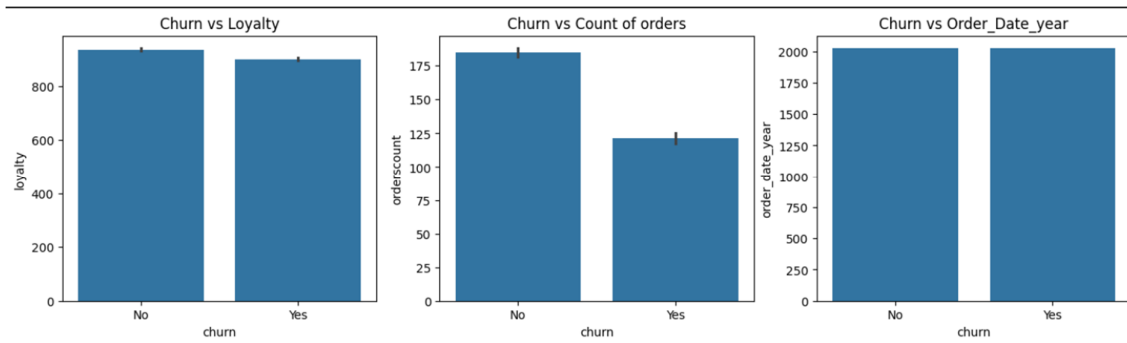
Exploratory Data Analysis (EDA)

Key Insights & Visualizations

- Top-selling books based on order count.
- Customer segmentation based on spending patterns.
- Seasonal demand trends for books.
- Churn rate analysis based on order frequency.
- Sample Visualization: (Include screenshots of charts and graphs.)

Churn predictor

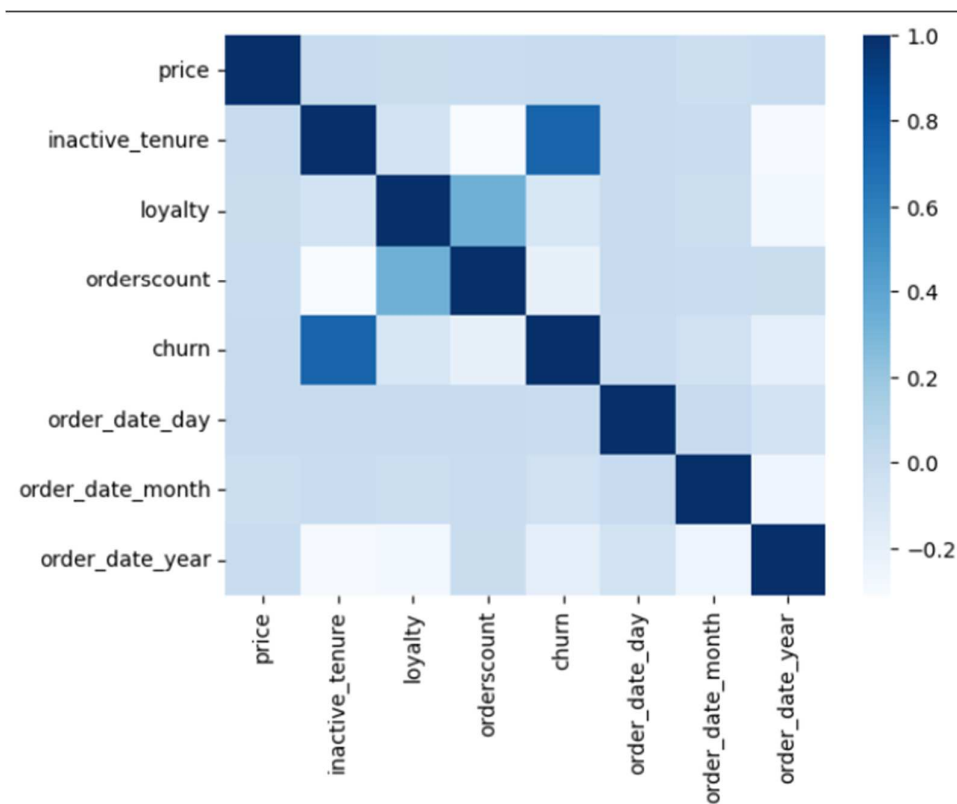




Observations:

- Higher the loyalty lower the churn People with less orders have churned
churn quality is evenly distributed for all years

Heat map of the correlation between numerical columns



Model Development

Customer Churn Prediction (ANN Model)

- Features Used: Total Orders, Total Spent, Average Order Value, Days Since Last Order.
- Model Architecture:
- Input Layer → Dense Layers (with LeakyReLU activation) → Output Layer.
- Evaluation Metrics: Accuracy, Precision, Recall, F1-score.

Demand Forecasting (LSTM Model)

- Features Used: Order count (with lag variables and moving averages).
- Model Architecture:
- LSTM Layers → Dense Layers → Output Layer.
- Evaluation Metrics: MAE, RMSE.

Model Deployment

Steps to Deploy on AWS EC2

1. EC2 Instance on AWS.
2. Transfer trained models using SCP command

```
scp -i ~/.ssh/bull.pem churn_model_multiclass.h5 ubuntu@<AWS_IP>:/home/ubuntu/
scp -i ~/.ssh/bull.pem lstm_demand_forecast.h5 ubuntu@<AWS_IP>:/home/ubuntu/
```

3. Install dependencies on EC2

```
sudo apt update && sudo apt install python3-pip
pip3 install tensorflow pandas streamlit joblib
```

4. Run Streamlit app on EC2

```
streamlit run app.py --server.port 8501
```

Streamlit Application

Features of the Dashboard

1. Customer Churn Prediction: Users input customer details to predict churn probability.
2. Demand Forecasting: Predicts book demand for upcoming months.
3. Personalized Book Recommendations:
4. Popular Books (based on highest orders).
5. Author-Based Recommendations (user inputs author name).

Challenges & Learnings

Challenges Faced

- Model input shape mismatches during inference.
- Deployment setup on AWS with permission errors.
- UI enhancements in Streamlit for better interaction.

Solutions Implemented

- Debugged input transformations to match model expectations.
- Ensured proper file transfer using SCP for deployment.
- Improved UI using Streamlit components & layout optimizations.

Conclusion & Future Scope

Summary of Key Findings

- Machine learning models provide valuable insights for the publishing industry.
- Churn prediction can help increase customer retention.
- Demand forecasting optimizes inventory management.

Future Improvements

- Enhance Book Recommendations using collaborative filtering.
- Incorporate NLP to analyze customer reviews for better insights.
- Deploy on cloud-based ML platforms for scalability.