4.2. Student Handout

Student Handout: Deploying the Trained Machine Learning Model in Power BI

Overview

This handout provides a concise guide on deploying machine learning models in Power BI, focusing on applying a trained model to a dataflow entity. By the end of this guide, you will understand the process of using a machine learning model to analyze new data and generate predictions.

Key Concepts

What is a Dataflow Entity?

- **Dataflow**: A collection of data from various sources (e.g., Excel files, databases, APIs) used for analysis in Power BI.
- **Dataflow Entity**: A specific table or dataset within a dataflow.

Examples:

1. A table of customer transactions imported from a CRM system.
2. A dataset of sales data from an Excel file.
3. A collection of product inventory data from an API.

Applying a Model to a Dataflow Entity

- **Scoring**: The process of using a trained machine learning model to analyze new data and generate predictions.

Examples:

1. Using a model to predict customer churn based on recent transaction data.
2. Applying a model to forecast sales for the next quarter using historical sales data.

3. Using a model to classify customer feedback as positive or negative.

Steps to Apply the Model

1. **Create a Dataflow**: Set up a dataflow in Power BI to gather data from various sources.
2. **Add a Dataflow Entity**: Import specific datasets into the dataflow.
3. **Train the Machine Learning Model**: Use historical data to train the model.
4. **Apply the Model to the Dataflow Entity**: Use the trained model to generate predictions on new data.

Examples:

1. Creating a dataflow for customer data and adding a table of recent transactions.
2. Training a model to predict product demand using past sales data.
3. Applying a model to classify incoming support tickets based on urgency.

Real-Time Scoring vs Batch Scoring

- **Real-Time Scoring**: Generating predictions as new data arrives.
- **Batch Scoring**: Generating predictions on a batch of data at once.

Examples:

1. Real-time scoring to predict purchase likelihood as customers browse an online store.
2. Batch scoring to identify potential churners from a daily customer activity log.
3. Real-time scoring to detect fraudulent transactions as they occur.

Handling Scoring Errors and Exceptions

- **Missing Data**: Impute missing values to ensure accurate predictions.
- **Outliers**: Remove or adjust outliers before applying the model.

Examples:

1. Filling missing customer age data with the average age before scoring.
2. Removing extreme sales values that skew predictions.

3. Adjusting outlier transaction amounts to improve fraud detection accuracy.

Monitoring Model Performance and Drift

- **Model Drift**: Occurs when patterns in new data change, reducing model accuracy.
- **Retraining**: Periodically update the model with recent data to maintain accuracy.

Examples:

1. Monitoring prediction accuracy for a customer churn model over time.
2. Retraining a sales forecasting model with quarterly data updates.
3. Adjusting a sentiment analysis model as language trends evolve.

Best Practices

1. **Regularly Monitor Model Performance**: Ensure the model remains accurate on new data.
2. **Retrain the Model Periodically**: Update the model with recent data.
3. **Handle Errors Gracefully**: Address missing data, outliers, and other issues.

Examples:

1. Setting up alerts for significant drops in model accuracy.
2. Scheduling monthly retraining sessions for a demand forecasting model.
3. Implementing data preprocessing steps to handle missing values before scoring.

Hands-On Example

1. **Create a Dataflow**: Set up a dataflow in Power BI.
2. **Add a Dataflow Entity**: Import customer transaction data.
3. **Train the Model**: Use historical data to train a churn prediction model.
4. **Apply the Model**: Generate predictions on new customer data.
5. **Monitor the Results**: Compare predictions to actual outcomes.

Conclusion

Applying a trained machine learning model to a dataflow entity in Power BI involves using the model to analyze new data and generate predictions. Regular monitoring and retraining are essential to maintain model accuracy.

Feel free to reach out with any questions or for further clarification on any of the topics covered in this handout.