EXPERNETIC

Data Science Internship – Assignment
Assessment Report

Shafa Bishirhafi

Table of Contents

1. Introduction

This report explains the implementation of the two tasks assigned as part of the data science and machine learning assessment.

1. **Task 1** focuses on utilizing machine learning and data analytics to extract valuable business insights from a dataset, with the goal of generating at least two insights that can drive business decision-making.
2. **Task 2** involves implementing a reinforcement learning solution, where an agent is trained to find the optimal path through a maze using Q-learning. This task aims to demonstrate how reinforcement learning can be applied to solve complex navigation problems.

Both tasks involve the application of key data science and machine learning. The following sections provide detailed explanations of the methodology, implementation, and insights for each task.

2. Task 1 - Implementing Machine learning models to generate valuable business insights

2.1 Objectives

The objective of Task 1 is to develop a machine learning and data analytics solution that can extract valuable business insights from a given dataset. This task aims to apply data analysis techniques to derive actionable insights for business improvement.

2.2 Methodology

1. **Dataset Overview and Preprocessing**
   The dataset consists of business-related data for 2 years. It was preprocessed to remove inconsistencies, handle missing values, and standardize the data for analysis.
   - **Data Cleaning:** Cleansing techniques were used to handle missing values, ensuring the dataset was complete and consistent, without any duplicates.
   - **Data Transformation:** The data was transformed into the correct datatypes and formats to enhance model development.
   - **Feature Engineering:** Relevant features were derived from the raw data to enhance model accuracy and interpretability.

2. **Model Selection**

    - **Model 1:** Apriori Algorithm

      An apriori algorithm was implemented to understand frequent buying patterns of customers to understand which two products are bought together which can help raise revenue for the business if it was considered. Moreover, they can also analyze the availability of promotions for those products.

    - **Model 2:** ARIMA model

      An ARIMA model was implemented to forecast the future sales for the top 5 products based on historical data. This model was trained to forecast future sales based on the weeks.

2.3 Business Insights

Model 1: Apriori Algorithm

- Cross-selling opportunities:

  Using the apriori algorithm, the business will get an understanding about the products frequently purchased together. Example: Products like spaghetti sauce and lasagna pasta tend to be bought together, meaning supermarkets can create bundled promotions or shelf placements for these items to increase sales.

- Store-specific insights:

  Store-specific insights can also be achieved because the business can get an understanding of different product associations in different supermarkets. This information can be used to personalize promotional strategies depending on store buying patterns, instead of a one-size-fits-all approach.

Model 2: ARIMA model

- Forecasting future demand:

  By using the sales forecast for the top 5 products businesses can use this information to identify if these top 5 products will continue to help increase revenue or will there be a downfall in the future. Based on this they can also estimate the number of units to produce to cater to the demand that may arise from future sales.

2.4 Conclusion

The two models implemented above will help the business achieve their goals immensely. These insights will be valuable for improving customer targeting and optimizing operational resources, which can have a significant impact on the business's performance.

3. Task 2 – Navigating a maze using Reinforcement learning

3.1 Objectives

The objective of Task 2 is to implement a reinforcement learning that helps the agent to learn to navigate a maze to reach a goal. The agent uses Q-learning, a reinforcement learning technique, to maximize its rewards and find the optimal path to reach the goal.

3.2 Methodology

1. **Defining the environment**
   Initially the environment for the maze was defined with 0 indicating free space and 1 representing a wall.

2. **Defining the action and rewards**
   The agent can perform 4 actions which is selected randomly – up, down, right, left
   The rewards were also defined as 1 for reward and -0.2 as a penalty.

3. **Q-Learning algorithm**
   The algorithm was implemented to train the model for a number of episodes and find the best route to reach the goal.

4. **Visualization**

   An animated visualization was created to view the movement from the agent's start position to the goal.

3.3 Conclusion

The implementation of the Q-learning algorithm in Task 2 was successful, with the agent consistently finding the optimal path in the maze. The agent's learning process was clearly demonstrated through visualization, showing how reinforcement learning can be applied to solve complex navigation problems.

4. Conclusion

This assignment demonstrates the use of machine learning and reinforcement learning techniques to solve real-world problems.

In **Task 1**, two valuable business insights—customer buying patterns and sales forecasting—were derived using data analytics and machine learning models. In **Task 2**, a reinforcement learning agent was successfully trained to navigate a maze using the Q-learning algorithm. Both tasks highlight the practical applications of data science and machine learning in driving business decisions and solving complex challenges.

The insights generated in Task 1 will help the business significantly to make informed decisions. Moreover, the skills gained from Task 2 can also help in applying to real-world scenarios.