HyperionDev

Machine Learning

Model Answer

Visit our website

Auto-graded task 1

1. For each of the following examples describe at least one possible input and output. Justify your answers:

1.1 A self-driving car

- **Input**: Real-time sensor data (cameras, LiDAR, radar) capturing the car's environment.
- **Output**: Control commands (steering, throttle, brake) for safe navigation.
- **Justification**: Sensor data informs the car's decisions, guiding its movements to navigate safely.

1.2 Netflix recommendation system

- **Input**: User data (viewing history, ratings, search queries) and content metadata.
- **Output**: Personalised content recommendations.
- **Justification**: User data informs the system to suggest content aligned with individual preferences.

1.3 Signature recognition

- **Input**: Scanned images or documents containing signatures.
- **Output**: Signature identification and verification accompanied by confidence scores.
- **Justification**: Scanned images are analysed to authenticate signatures against reference samples.

1.4 Medical diagnosis

- **Input**: Patient data (medical history, vital signs, symptoms, test results).
- **Output**: Diagnosis or prognosis for treatment planning.
- **Justification**: Patient data guides healthcare professionals in diagnosing and managing medical conditions.

2. For each of the following case studies, determine whether it is appropriate to utilise regression or classification machine learning algorithms. Justify your answers:

2.1 Classifying emails as promotion or social based on their content and metadata.

- Classification: It involves categorising emails into distinct classes.

Copyright © 2024 HyperionDev. All rights reserved.

2.2 Forecasting the stock price of a company based on historical data and market trends.

- Regression: This task entails predicting a numerical value (the price) based on historical data.

2.3 Sorting images of animals into different species based on their visual features.

- Classification: The machine learning (ML) model categorises species based on their visual features.

2.4 Predicting the likelihood of a patient having a particular disease based on medical history and diagnostic test results.

- Regression: It involves predicting likelihood, which is a continuous data type. However, if the outcome is binary, that is, whether a patient has a disease or not, it would be a classification problem

3. For each of the following real-world problems, determine whether it is appropriate to utilise a supervised or unsupervised machine learning algorithm. Justify your answers:

3.1 Detecting anomalies in a manufacturing process using sensor data without prior knowledge of specific anomaly patterns.

- Unsupervised: The ML algorithm autonomously detects anomalies in manufacturing processes without specific anomaly patterns.

3.2 Predicting customer lifetime value based on historical transaction data and customer demographics.

- Supervised: Historical transaction data and customer demographics are used as labelled data to train the model for predicting future customer lifetime values.

3.3 Segmenting customer demographics based on their purchase history, browsing behaviour, and preferences.

- Unsupervised: Unsupervised learning segments customer demographics by autonomously identifying patterns and grouping customers with similar behaviours and preferences.

3.4 Analysing social media posts to categorise them into different themes.

- Supervised: Supervised learning categorises social media posts into predefined themes using labelled data, enabling accurate classification. However, this could also be unsupervised learning if the data is not labelled.

Copyright © 2024 HyperionDev. All rights reserved.

4. For each of the following real-world problems, determine whether it is appropriate to utilise semi-supervised machine learning algorithms. Justify your answers:

4.1 Predicting fraudulent financial transactions using a dataset where most transactions are labelled as fraudulent or legitimate.

- Not appropriate. Since most transactions are labelled, it is more suitable to use supervised learning techniques rather than semi-supervised learning.

4.2 Analysing customer satisfaction surveys where only a small portion of the data is labelled with satisfaction ratings.

- Appropriate. Since the labelled data is scarce, the model can use labelled data as a guide and unlabeled data to capture the underlying structure of the data, enhancing the model's performance.

4.3 Identifying spam emails in a dataset where the majority of emails are labelled.

- Not Appropriate. Similar to 4.1, since the majority of emails are labelled, supervised learning techniques would be more appropriate.

4.4 Predicting the probability of default for credit card applicants based on their complete financial and credit-related information.

- Appropriate. Semi-supervised learning could be suitable here, especially if there is a large amount of unlabeled data available. Combining labelled (applicants with known default status) and unlabeled (applicants without known default status) data can enhance predictive performance. If the input data is fully labelled(default or no default), supervised learning would be more suitable.

Copyright © 2024 HyperionDev. All rights reserved.