ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING - 6CS012

COURSEWORK 3:

QUESTION AND ANSWER

Student Name: Saroj Devkota

University ID: 232955

Group: L6CG6

Module Leader: Mr. Siman Giri

Module Tutor: Mr. Shiv Kumar Yadav

Submission Date: 5/14/2025

ABSTRACT

This coursework investigates the application of machine learning techniques in Nepal’s digital payment industry, focusing on eSewa. It proposes unsupervised models like DBSCAN, Hierarchical Clustering, PCA, and t-SNE for analysing service usage trends across regions and customer groups, supporting market expansion strategies. It also explores unsupervised anomaly detection methods, including Isolation Forest, Autoencoders, and LOF, for identifying fraudulent transactions without labelled data. Additionally, the document covers key deep learning concepts such as overfitting and its solutions and compares CNNs and RNNs based on their suitability for image-based and sequential data, offering a balanced practical and theoretical guide for data-driven fintech solutions in Nepal.

TABLE OF CONTENTS

1. LONG QUESTION

You are a Data Scientist at eSewa, Nepal’s leading digital payment platform.

Answer:

1. SERVICE USAGE TREND ANALYSIS FOR MARKET EXPANSION STRATEGY

PROBLEM STATEMENT

eSewa diversifies its portfolio adding such services as digital loans, QR payments at restaurants, and interbank transfers. Yet, there is no systematic means of evaluating the service categories that are popular in certain geographical territories and among certain customer segments. The lack of labelled adoption trends or predefined success indicators for services makes it difficult to detect emerging patterns of use that could be used to guide a strategic business decision.

PROPOSED APPROACH

To uncover underlying patterns in service usage, this study proposes the application of unsupervised machine learning techniques, including DBSCAN, Hierarchical Clustering, and Dimensionality Reduction methods such as PCA and t-SNE. These techniques would be applied to multi-dimensional datasets encompassing:

- Transaction volumes per service type
- Temporal growth trends (e.g., monthly or quarterly usage rates)
- Region-wise or location-specific activity
- User demographics, including age, gender, and profession (where available)
- Such analysis would help reveal natural clusters or trends, identifying which services are popular among certain user profiles or regions, without relying on labelled data.

BUSINESS INTEGRATION

- Insights derived from this analysis can support strategic decision-making across various domains, including:
- Deployment of hyper-localized marketing campaigns based on regional demand
- Prioritization of high-demand services in areas with rapid adoption
- Identification of underperforming services for targeted improvement or replacement
- Planning of regional merchant onboarding drives, particularly for QR payments or remittance services in growth regions.

By leveraging these data-driven insights, eSewa can adopt a proactive and scalable approach to market expansion, ensuring that user adoption trends guide business priorities and resource allocation.

2. PATTERN FOR FRAUD DETECTION AND CLUSTERING OF ANOMALIES

PROBLEM STATEMENT:

As we begin to see the increase in digital transactions, along comes the increase of threat to fraudulence, starting from fake merchant transactions, unauthorized account access, to money laundering schemes. Traditional fraud detection systems will commonly use supervised learning models built off historical data on fraud, but fraudsters are always looking for new ways of fraud. Furthermore, a lot of fraudulent activities are not detected because of the absence of labelled data.

PROPOSED APPROACH:

Unsupervised anomaly detection methods such as Isolation Forest, Autoencoders, and Local Outlier Factor (LOF) can be used to transaction data (amount, frequency, time, location, device, and recipient patterns) in order to identify aberrant behaviours. Clustering algorithms like DBSCAN can help to find the groups of transactions or users that behave oddly with the help of unlabelled fraud cases.

BUSINESS INTEGRATION:

An unsupervised fraud detection system would work as an early warning system and identify the transactions or users that look suspicious for a manual review. It can minimize costs, save the company image, and enhance a customer’s trust. Furthermore, it supports current rule-based or supervised models by discovering the hidden fraud patterns.

2. SHORT QUESTIONS

2.1 OVERFITTING IS A COMMON CHALLENGE IN DL MODELS

Answer: Overfitting is a very common phenomenon in the deep learning whereby a model performs extremely well on its training set data, but fails dismally when presented to new, never before seen data. This occurs because the model becomes memorizing particular facts and noises of the training set rather than learning the general patterns it is designed to detect. To address this, two common techniques are commonly applied: **Dropout and Early Stopping:**

- What **Dropout** does is that it turns off a random number of neurons during every training step. This introduces certain randomicity which makes the model less reliant on particular neurons or path. Consequently, the model learns to discriminate more trustworthy general patterns. For instance, in an image classification for an e-commerce site, a dropout rate of 0.5 put on the dense layers can avoid the model from overemphasizing the useless details such as unique lighting or pixel arrangements.
- **Another good approach is Early Stopping:** It monitors the performance of the model on a different set of validation data while it is being trained. If the model's ability to detect the true answers on this set of validation data does not continue to improve or even starts getting worse after a few training replays (called epochs), the training process is stopped. This allows the model not to get too adaptive to the training set. For example, in a sentiment analysis where feedback is dichotomized as positive and negative, early stopping may stop the training if not improved for 10 successive epochs.

Both of these methods assist deep learning models to do better on new unseen data. In their combination, they substantially decrease the overfitting and thus make models more useful and trustworthy in the real world.

2.2 DIFFERENCE BETWEEN CNN AND RNN

Answer: Convolutional neural networks (CNNs) are customized in a way that they can process structured data such as images. Their grid-like structure agrees with the structure of image data rendering them most suitable for discovering patterns similar to edges, corners, and textures. CNNs apply several filters (or kernels) for scanning the image where these patterns are automatically learnt without the need to extract features by hand. This is what makes CNNs very powerful for various computer vision jobs including image classification, face recognition, even product matching on e-commerce sites.

Recurrent Neural Networks (RNNs) are designed to handle data in which the sequence or order of pieces of data is a concern. They involve a type of memory, which enables them to use information from the past steps to enhance current predictions. This is why RNNs are suitable for such tasks as forecasting the price of stocks based on the history of their price change, analysis of the patterns of customer purchase, or real-time language translation.

The major difference between CNNs and RNNs is that CNNs are elements that are aimed at learning a certain kind of structure, whereas RNNs stand for a set of components to learn any type of structure. The CNNs are concerned with spatial relation the placement and connectivity of pixels within an image. RNNs are on the other hand concerned with temporal relationships how information alters in the time.

As strong as it is, training deep learning models is not without its problematic aspect. One of them is the vanishing gradient problem, in particular, in long or deep networks. This happens when gradients get too small to efficiently update during training resulting in slowing or halting of learning. Another of the other major challenges is the problem of overfitting whereby a model performs outstandingly on the training data set but cannot generalize in unseen test data.

To address such issues, approaches such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) are employed in RNNs in order to have an improved ability to perform long term dependencies as well as retain information of value over time. In CNNs, batch normalization is commonly used with the purpose of stabilizing and improving training. In addition, similar approaches are used by both types of networks, namely the dropout, regularization and early stopping and their purpose is to avoid overfitting and achieve better real-world performance.