Project Report: Understanding Artificial Intelligence

Danny Xiao Yong Wu
202532469

Component One

1. Introduction

Streaming services collect a tremendous amount of customer data for analyzing user behavior, spending trends, and subscription decisions. Evaluating how much customers spend and identifying users at risk of churn are critical to maximizing customer retention and optimizing the strategies of your business in the ever-competitive digital market by segmenting customers into meaningful segments.

This study aims to explore how supervised and unsupervised machine learning algorithms learn from a streaming service dataset. More specifically, a regression model is created to estimate Customer Monthly Spend, a classification model predicts whether customers will churn, and clustering is used to identify distinct customer segments based on their behavior.

Simple and multiple regression models are first presented using numerical features, and then models using numerical and categorical features are applied. Next, more complex models such as a Random Forest Regressor and an Artificial Neural Network are implemented. In addition, classification models are trained and tested for churn prediction, and unsupervised learning methods are applied for customer segmentation. The models are tested using appropriate metrics along with charts in support of the findings.

2. Data Set and Methodology

In the context of data collection, the dataset used in this study comprises data on streaming service customers, including numerical and categorical characteristics of user behavior, subscription attributes, and spending behavior. The target variables for regression analysis are Monthly Spend and Churn for classification tasks, with the remaining features used as input variables.

Data was preprocessed before analysis; thus, training models was conducted on the dataset. Missing values were verified and handled as required, and categorical values were converted to numerical values. Numerical features were scaled if the model required it—especially in models sensitive to feature magnitude, such as clustering algorithms and Artificial Neural Networks.

The dataset was split into training and testing sets to facilitate an unbiased assessment of model performance. Regression models were evaluated using metrics such as the coefficient of determination (R^2) and root mean squared error (RMSE). The classification model was rated on the basis of accuracy, precision, recall, F1 score, and confusion matrices. Clustering techniques were analyzed using clustering metrics and visual charts.

This enables reliable and interpretable results, allowing comparisons among different machine learning models without sacrificing consistency in the outputs of their use.

3. Regression Models

In this particular subsection several regression models are described; these are intended to predict Customer Monthly Spend. A series of progressive regression methodologies has been chosen, with simplified regression models being applied starting from simple regression models and increasing the model complexity step by step. This would enable us to comprehensively assess how performance of additional features and advanced learning methods is reflected in predictive accuracy. We used the same quantitative evaluation metrics for measuring the evaluation across all models to allow fair and objective comparisons.

3.1. Single-Feature Regression

We produced the simplest single-feature regression models using a single numeric variable to predict Monthly Spend. In order to explore the relationship between each feature and customer spending behaviour, both linear and nonlinear regression methods have been tested. Model performance was evaluated as a function of the coefficient of determination (R^2) and root mean squared error (RMSE), whereby both fit and prediction error could be evaluated.

The results showed that while some numerical features were significantly linked to Monthly Spend, the overall predictive power of single-feature models was fairly weak. At least some nonlinear models were better than linear models in several instances due to capturing the curves or complexity of the relationships of the feature and spending. This implies that the consumer's spending behaviour is not merely linear. However, even with this development, single-feature models are not very useful as explanations to the customers alone, because they do not explain jointly the combination of the behaviour- and subscription-related effects for customer spending.

The findings on spending behaviour highlight the limitations of individual attribute-driven, single customer predictions. Although single-feature models may offer an initial perspective on relationship between variable and Monthly Spend, they do not model interactions between features. Thus, they are mostly exploratory rather than predictive, because they build a benchmark for evaluation with more complex models.

3.2. Model of Multiple-Feature Feature Predictive Model

We trained regression models with multiple numerical features to improve predictive capability. Since the objective of these models is to integrate multiple variables, the model would prefer the input features and describe what underpins customers' consumption. Compared to single-feature regression, these models had higher R^2 and lower RMSE values, indicating a distinct increase in prediction capabilities. These results suggest that the influence of multiple factors and components, and not a single variable, dominates the Monthly Spend.

Although it is more interpretable and better fit, multivariate regression is not perfect and is still heavily limited by classical regression assumptions. Especially if the linear model is too small or complex, the nonlinear relationships and the complex relationships among parameters are not fully captured. This encourages more sophisticated machine learning models.

For customers in the middle of their spend behaviour, the incremental improvement achieved by multiple-feature regression implies that customer spending behaviour is determined by multiple interaction variables of behaviour and subscription. Nonetheless, classical regression models are linear and cannot easily account for nonlinear dependencies and feature interactions. In the face of this limitation, we are calling for more flexible machine learning models in the context of complex patterns in customer behaviour.

3.3. Regression with Categorical and Numerical Features

For this feature inclusion of numerical and categorical variables, a Random Forest Regressor was used. This model is good for this purpose since it can perform the natural mixing of data types, and it does not require explicit feature transformation to observe nonlinear relationships. Prediction error and explanatory power were better with the Random Forest Regressor than with conventional linear regression methods. This enhancement emphasises the need to look at categorical features, including customer attributes and subscription characteristics, for modelling spending behaviour.

Moreover, feature importance analysis was also performed, offering insight into which variables most decisively have a major effect on Monthly Spend. This adds additional interpretability for the model while still preserving high predictive accuracy.

Besides predictive gains, the Random Forest model also has real-world applications. Its robustness to noise and reduced sensitivity to outliers make the model very important for customer data that typically exhibits a lot of variability and erratic response. Moreover, the ensemble-based structure of the model decreases overfitting, so that more robustness can be achieved for any data set that has never been observed.

3.4. Artificial Neural Network (ANN)

The machine learning model ANN was trained to model the more complicated data relationships. The ANN adopts a structure of multi-layered networks and non-linear activation functions hence can learn very complex features that are not easily understood and processed by classical regression methods.

The ANN achieved predictive efficacy like the Random Forest Regressor, thereby proving its robustness in modelling nonlinear behaviour. But this performance still required more computational resources as well as careful hyperparameter tuning. Therefore, the ANN was able to offer only slight improvements compared to its complexity, suggesting that ensemble-based, simpler models may be a more effective solution for this dataset.

3.5. Best Regression Model

The Random Forest Regressor achieved the best overall performance across evaluation metrics. With the lowest prediction error and strong interpretability, it performed best for predicting Monthly Spend.

4. Churn Prediction

The classification models were trained to predict if a customer will churn (1) or not (0). The performance of the models was measured with accuracy, precision, recall, F1-score, and confusion matrices. These metrics present a balanced perspective, especially if a churn prediction requires identifying customer groups on the verge of churn to be more critical than overall accuracy.

Out of all the models tested, ensemble-based classifiers produced superior recall and F1-scores over other methods. This is where high recall is crucial because it tells the model that you will be able to predict who is likely to churn. This makes these models very useful for helping with proactive retention efforts to minimize customer loss.

5. Segmentation

Unsupervised learning techniques were employed for characterizing customer segments based on behavioural and subscription attributes. First, the k-means algorithm was used, calculating an optimal cluster number using the elbow method and the silhouette score calculation. This guaranteed that a sufficient number of clusters kept both the cohesion within clusters and the separation between clusters.

The resulting clusters provided significant differences in consumers' behavioural patterns, especially in terms of spending behaviours and engagement patterns. The segmentation can serve as separate customer profiles that could lead to specialized marketing efforts and services.

Another clustering method was also employed for comparison. Although clustering results yielded different segment structures, k-means produced the most stable and interpretable results when accompanied by clustering metrics and visual analysis.

These customer clusters represent actionable insights that can guide marketing strategies, pricing plans, and engagement activities. For instance, high-spending but low-engagement groups could use focused incentives, whereas low-spending but highly engaged users could present opportunities for upselling. This type of segmentation allows more individualised decision-making as opposed to treating all customers similarly.

6. Conclusion

Customer behaviour in streaming services can be properly analyzed using supervised and unsupervised learning methods. Models that capture non-linear relationships and categorical information outperform simpler approaches, while classification and clustering techniques successfully identify churn risk and meaningful customer segments.

In general, ensemble-based models achieve the optimal trade-off between accuracy and interpretability. The results highlight the importance of combining multiple analytical approaches to gain a comprehensive understanding of customer behaviour. Future work may further improve performance by incorporating larger datasets, additional features, or alternative modelling techniques.

Component Two

1. Introduction

Assessment of vehicle damage is an essential part of the insurance claim verification process that directly influences the speed and reliability of claim decisions. This is a traditionally manual inspection task that is time-consuming, subjective, and prone to human error. As more and more image data are available, the automated image-based classification of damage has emerged as a relevant area of application for machine learning in the insurance sector.

Here I describe the performance of the Convolutional Neural Network (CNN) to classify different forms of vehicle damage from images. Crack, scratch, dent, tire flat, glass shatter, and lamp broken are the classified damages. This task is particularly well-suited for CNNs given their ability to extract spatial features automatically from the images and learn complex visual patterns. The model is trained using the Vehicle Damage Insurance Verification dataset from Kaggle, with the objective of accurately identifying the types of damage in order to provide insurance claim validation value.

Here, I analyse that the CNN architecture was designed, regularisation techniques have been used to enhance generalisation, hyperparameter tuning and its influence on the performance of the models. Also, the training and validation results are examined to detect signs that may represent overfitting, reinforced by appropriate performance metrics and visualisations.

2. Dataset and Preprocessing

The dataset being used for this work is a dataset for Vehicle Damage Insurance Verification taken from Kaggle. It includes annotated pictures depicting the various forms of vehicle damage, including crack, scratch, dent, tire flat, glass shatter, and broken lamp. Each image is assigned to one damage type, making this a multi-class image classification problem.

Prior to the Convolutional Neural Network being trained, some preprocessing was done to make sure the data was acceptable for data fit towards model training. The images were all rescaled to the same resolution in order to be used as an equal number for operation by the CNN architecture. Pixel values were normalised to induce numerical stability during training and help the model converge more efficiently.

Training and validation data sets were separated from each other to ensure that model performance on new data is evaluated objectively. This separation is crucial to detect overfitting and evaluate the model’s generalisation abilities outside of the training sets. As is widely done during training, where possible, data generators were applied to load images in batches efficiently.

As a result, these preprocessing steps lead to homogeneous, well-scaled, and well-structured input data suitable for conducting convolutional learning, serving as an effective basis for evaluating the CNN architecture and later modelling decisions.

3. CNN Architecture (Part A)

A CNN (Convolutional Neural Network) was constructed using a sequential architecture to classify the images into six different categories of damage: crack, scratch, dent, tire flat, glass shatter, and lamp broken. To the network, a colour image with dimensions 256 × 256 × 3 is input, which helps the model learn not by using human features that were carefully designed, but from the spatial structure of the damage in the image.

For this purpose, a baseline CNN is built with a basic feature extraction → classification structure in mind. During feature separation, with three convolutional blocks using an increasing number of filters, it aims to classify more and more sophisticated visual cues. Namely, the network applies the following:

- Conv2D (32 filters, 3×3, ReLU) → MaxPooling (2×2)
- Conv2D (64 filters, 3×3, ReLU) → MaxPooling (2×2)
- Conv2D (128 filters, 3×3, ReLU) → MaxPooling (2×2)

Small 3×3 kernels help the model to learn about local texture changes such as scratches or cracks, while increasing blocks through multiple convolutional layers helps the model to learn higher-level features as the depth grows. The feature maps are spatially reduced by pooling layers, which reduce computation and improve the model's ability to capture the most informative patterns.

The feature maps obtained after the convolutional blocks are then transformed into a one-dimensional representation using a Flatten layer. The classification phase subsequently employs two fully connected layers:

- Dense (128 units, ReLU)
- Dense (64 units, ReLU)

Finally, the output layer is:

- Dense (6 units, Softmax)

The Softmax result provides the probability distribution for the six damage classes, which can be used for multi-class classification. The model was trained using the Adam optimiser (baseline learning rate 0.001) and the categorical cross-entropy loss function, in line with one-hot encoded class vectors used to apply the labels.

The model performance was monitored through measures of accuracy, and the architecture was subsequently expanded and verified through further design decisions (e.g., dropout, batch normalisation, and pooling tuning) to test how structural changes alter validation performance and generalisation.

4. Regularisation Methods (Part B)

To address this issue some regularisation was introduced to the CNN architecture to optimise generalisation ability and prevent overfitting. Considering the relatively small size of the image dataset and the complexity of convolutional-based models, regularisation is important for preventing the network from memorising training samples rather than learning transferable visual features.

An important regularisation method used was dropout, which was added after the fully connected layers. Dropout randomly deactivates a subset of neurons during training, which decreases dependence on specific activations and stimulates the model to learn more informative feature representations. Applying dropout resulted in a marked decrease in the difference between training and validation accuracy, denoting an enhanced generalisation.

Additionally, batch normalisation was performed in certain model variants. In this way the intermediate layer activations become normalised during training, stabilising the learning and speeding the time to convergence. In addition, batch normalisation helps to regularise the training by lowering the sensitivity to weight initialisation and suppressing internal covariate shift.

The effect of these techniques was judged by examining training and validation accuracy and loss curves. Models with regularisation showed smoother learning behaviour and lower overfitting than the baseline CNN. Although training accuracy could sometimes be a bit lower, validation performance was more stable, demonstrating improved generalisation to images not previously seen.

Overall, the dropout and batch normalisation effectively limited overfitting while achieving promising classification results.

5. Hyperparameter Tuning (Part C)

When performing hyperparameter tuning to analyse different changes to the training and architecture settings that affected the CNN performance. Various configurations were performed to systematically test those with validation performances that benefited from stable learning behaviour while simultaneously improving validation performance.

The learning rate was analysed as a crucial parameter given its effect on the speed of convergence and the stability of the training. Small learning rates resulted in slower, more stable learning; larger values led to fluctuating validation performance. Accordingly, the Adam optimiser with the learning rate = 0.001 was chosen, which strikes a great balance of accuracy and speed of convergence.

The batch size was also evaluated for its impact on training dynamics. As batch sizes were smaller, gradient updates became more variable and improved generalisation but increased training time. Larger batch sizes resulted in smoother training curves but less validation accuracy. A moderate batch size was thus selected to find a compromise between efficiency and performance.

The number of epochs was adjusted to be based on the validation loss and accuracy and on the principle of overfitting due to overfitting and underfitting due to too little. The number of epochs was picked out for optimising accuracy of validations. Architectural properties including the filter number and pooling algorithms were also tested, with higher filters increasing the number of features extracted, and diminishing returns being observed. The validation with the max pooling was best performance.

Hyperparameter tuning, on the whole, showed that the general performance of a CNN is strongly conditional on configuration of the training, although with good choice of parameters one has to achieve high accuracy and generalisation.

6. Overfitting Analysis (Part D)

Overfitting analysis was carried out to test generalisation capability of the CNN to unseen data and evaluate the effectiveness of the applied regularisation and hyperparameter tuning. Training and validation accuracy and loss curves were studied in various model settings.

In the baseline CNN, obvious difference in performance with relation to training and validation was demonstrated after several training iterations. As the training accuracy increased, the validation accuracy plateaued or fluctuated, which suggests that the model was memorising training data rather than learning generalisable features. This was reinforced by increasing validation loss when training loss reduced.

This gap was reduced when regularisation methods like dropout and batch normalisation were employed. Validation accuracy became more consistent and validation loss trended more smoothly which signifies that the model was more generalised and less sensitive to training noise. Hyperparameter tuning also enhanced the control of overfitting through a well chosen learning rate and number of epochs.

On the whole, the final model had a balanced learning behaviour, demonstrating good validation results and low overfitting.

7. Conclusion

This study shows that Convolutional Neural Networks are suitable for the task of vehicle damage classification using image data. As the proposed CNN architecture learned useful visual features specific to different types of vehicle damage, reliable classification performance was achieved across several damage categories.

Due to the use of methods such as dropout and batch normalisation, the model gained an important improvement in generalising to unseen data. Moreover, in terms of training stability and validation performance, hyperparameter tuning was found to play an important role, demonstrating the necessity of choosing proper learning rate, batch size, and architectural parameters. The overfitting findings further confirmed that these strategies control model complexity and reduce the gap between the training set and the validation set.

The final CNN model shows a fine trade-off between predictive accuracy and generalisation. While the above-mentioned architectures could learn complex patterns, standardised and simpler regularised models often yielded comparable performance with greater efficiency. In future work, similar analyses could be performed using larger datasets, transfer learning methodologies, or additional architectures, which may help improve classification performance and further enhance generalisation robustness.

Component Three

Ethics of AI in Transportation

1. Introduction

Artificial Intelligence (AI) is being used to a greater extent in the transportation industry, especially autonomous and semi-autonomous vehicles, in order to enhance safety, efficiency, and mobility. Although AI systems assist in perception, decision-making, navigation and risk assessment, their increasing autonomy raises pressing ethical issues associated with transparency, fairness, trust, and accountability. The choices made by autonomous vehicle algorithms can have significant human safety impact, making ethical considerations very important in the development and implementation of such AI.

This report offers a critical look at ethical implications of AI in transportation under autonomous vehicles. Based on peer-reviewed literature, it illustrates the main technical challenges in Explainable AI, Fair AI, and Trustworthy AI, examines the existing ethical frameworks, and provides technical and organisational solutions to support a responsible deployment of AI in transportation.

2. Technical Challenges in Ethical Transportation with AI

2.1. Explainable and Transparent AI - (Si & Araz Taeihagh, 2019)

One of the notable technological issues around transportation AI is the lack of transparency and explainability in autonomous vehicle decision-making. Deep learning models, especially convolutional and reinforcement learning networks, constitute one of the most important “black boxes” for all manner of autonomous vehicles employed now. These models have high predictive power and accuracy, but are very opaque to engineers, regulators, and end users.

Lim and Taeihagh (2019) hypothesize that such opacity poses ethical and safety issues because it becomes difficult to understand or justify the decision that an autonomous vehicle made in a certain critical context. For instance, the failure to articulate the reasoning behind an algorithm in the context of a crash or near one makes investigating accidents and imposing liability on algorithmic actors difficult in the case of incidents, and it erodes public confidence. Explainable AI techniques are extremely difficult to deploy in live driving with high complexity autonomy and AI-based decision speed.

2.2. Fairness and Bias in Autonomous Vehicles - (Si & Araz Taeihagh, 2019)

Fair AI is yet another big ethical challenge to transportation systems. Training data, sensor limitations, or design assumptions may bias driver-less vehicles. If perception models are trained on non-representative datasets of certain environments, road users, or demographic groups, then there is potential for biased performance of the models across different populations.

Lim and Taeihagh (2019) also find that autonomous vehicles might be less reliable at detecting pedestrians or cyclists under specific lighting conditions, urban environments, or certain geographical regions. These disparities have important ethical implications as differential performance can increase the risk to some groups. Technically, reducing bias is hard as it requires diverse datasets, continuous monitoring, and fairness-aware evaluation metrics, which increases system complexity and development costs.

2.3. Trustworthy AI and Public Acceptance - (Naiseh et al., 2024)

AI must be trustworthy for the implementation of widespread autonomous transportation systems. Trust is not only about the technical reliability but also the perception of the users of safety, fairness, and accountability. Naiseh et al. (2024) demonstrate that public trust in autonomous vehicles is significantly influenced by perceived risk, system transparency, and prior experience with AI technologies.

On the technical side, trustworthy AI means having robustness, reliability, and fail-safe behaviour under uncertain conditions. However, high reliability in all scenarios is still a big hurdle ahead of us. Edge cases (anomalous road conditions, deviation in human behaviour), on the other hand, are difficult to predict and model, seriously compromising trust in fully autonomous systems.

3. Ethical Framework Evaluation - (Rhim et al., 2021) (Naiseh et al., 2024)

There are multiple ethical frameworks available that have been suggested to inform the responsible construction and implementation of AI in transportation. These frameworks aim to address challenges related to transparency, fairness, accountability, and safety in the design of transport.

Rhim et al. (2021) present an ethical decision-making framework for autonomous vehicles that incorporates moral reasoning with algorithmic design. One positive aspect of this approach is its focus on formalising ethical principles for technical systems to make more uniform ethical decision making for morally complex situations. But the model is limited by its inability to transfer its moral values to the real-world environment as these are culturally-bound and hard to encode universally.

Similarly, the trust-oriented frameworks highlighted by Naiseh et al. (2024) focus on human-centered design and risk communication. Such frameworks give robust insights into user acceptance and transparency but frequently do not present concrete technical guidance on practical implementation. As such, they may be either abstract or in some way hard to operationalize in complex autonomous vehicle systems.

Overall, existing ethical frameworks do offer important high-level guidance, yet still find it challenging to connect the dots between ethical principles and actual system design. Real-world transportation settings and their constraints — computation, regulation, and economic viability are not always fully accounted for in theoretical models.

4. Innovation and Potential Solutions - (Si & Araz Taeihagh, 2019) (Naiseh et al., 2024)

The practical use of AI in transportation should improve ethically through technical innovation and organisational approaches. One such method is the use of Explainable AI layers in autonomous systems. Post-hoc explanation tools can clarify model decisions after critical events, and help us in maintaining real-time performance while building accountability.

Another crucial development is the utilization of human-in-the-loop systems, especially at initial stages of deployment. Enabling human intervention where it is required, in high-risk circumstances, strikes that delicate balance between autonomy and ethical duty and fosters trust through human accountability.

In terms of fairness, bias auditing pipelines should be included at every stage in the system life cycle to achieve continuous monitoring across environments and user groups. These pipelines enable discrepancies to be identified and addressed, thus reinforcing fairness as a process that goes on rather than a one-time design goal (Si & Araz Taeihagh, 2019).

Lastly, mechanisms for regulatory and transparency are critical. Having such documentation about limitations, decision boundaries, and failure cases should help meet legal and privacy standards; more importantly, it would build public trust and help meet compliance requirements. Therefore, ethical deployment requires collaboration among engineers, policymakers, and social stakeholders.

5. Conclusion

While AI applications in transportation can be highly beneficial, they can lead to fundamental ethical challenges around issues of explainability, fairness, and trustworthiness. These issues are related to technical design choices and cannot be addressed through performance improvements alone. Previous research notes that ethical frameworks can offer guidance but implementation remains complex due to technical and social constraints.

The report finds that ethical AI in transportation necessitates transparent models, fairness-aware design, robust validation, and appropriate human oversight. Integrating explainability mechanisms and trust-centred design practices is essential to ensure that autonomous technologies develop in a manner that aligns with societal values and public safety.

Reviewed Articles

Si, H., & Araz Taeihagh. (2019). Algorithmic Decision-Making in AVs: Understanding Ethical and Technical Concerns for Smart Cities. *Sustainability*, *11*(20), 5791–5791. https://doi.org/10.3390/su11205791

Rhim, J., Lee, J.-H., Chen, M., & Lim, A. (2021). A Deeper Look at Autonomous Vehicle Ethics: An Integrative Ethical Decision-Making Framework to Explain Moral Pluralism. *Frontiers in Robotics and AI*, *8*. https://doi.org/10.3389/frobt.2021.632394

Naiseh, M., Clark, J., Akarsu, T., Hanoch, Y., Brito, M., Wald, M., Webster, T., & Shukla, P. (2024). Trust, risk perception, and intention to use autonomous vehicles: an interdisciplinary bibliometric review. *AI & SOCIETY*, *40*(2), 1091–1111. https://doi.org/10.1007/s00146-024-01895-2

Fernández Llorca, D., Hamon, R., Junklewitz, H., Grosse, K., Kunze, L., Seiniger, P., Swaim, R., Reed, N., Alahi, A., Gómez, E., Sánchez, I., & Kriston, A. (2025). Testing autonomous vehicles and AI: perspectives and challenges from cybersecurity, transparency, robustness and fairness. *European Transport Research Review*, *17*(1). https://doi.org/10.1186/s12544-025-00732-x