

**Asia Pacific University of Technology & Innovation,**

**Jalan Technology Park Malaysia**

**DATA MANAGEMENT ASSIGNMENT PART 1: FEATURE ENGINEERING**

**INTAKE CODE: APDMF2112DSBA(DE)(PR)**

**MODULE CODE: CT051-3-M-DM**

**HAND OUT DATE: 01 March 2022**

**HAND IN DATE: 15 April 2022**

**LECTURE NAME: Assoc. Prof. Dr. Raja Rajeswari**

**STUDENT NAME: Linda Houmed Bililis**

**TP NUMBER: TP060864**

**TABLE OF CONTENT**

[**I.** **INTRODUCTION** 4](#_Toc100955721)

[**1.** **Feature Transformation** 4](#_Toc100955722)

[a) Recode Range 5](#_Toc100955723)

[**2.** **Feature Creation** 5](#_Toc100955724)

[**Conclusion** 5](#_Toc100955725)

[**REFERENCES** 6](#_Toc100955726)

**Abstract**

Feature engineering is outlined as a technique of selecting, manipulating, and reworking datasets into features that will be used in supervised learning. It is an important undertaking to prepare a dataset for machine learning. Hence, it requires cleaning data before training the model. Hence data preprocessing is conducted which is an imputation of missing values, elimination of outliers, data exploration, and so on. feature engineering is dived into two: first, feature transformation, which is the method of modifying the structure of data. With log transformation, the distribution of data is converted into the normal distribution. The feature transformation is used to identify whether the data is a classification problem or regression. Then feature creation, which is a method of creating a new feature from the existing feature. This feature may be more efficient to perform a machine learning model. In this, paper log transformation and standardization are used in part 2 and recoding the target variable is used as well. However, the research doesn’t involve a feature creation method.

# **INTRODUCTION**

Feature engineering is defined as a method of choosing, manipulating, and remodeling datasets into features that will be utilized in supervised learning. According to Nargesian et al. ( 2017), Feature engineering could be a central task in data preparation for machine learning. It is the application of constructing an appropriate feature from a given feature that causes improved prophetical performance. Therefore, machine learning models are created to be strained as a result of such constraints are crucial to the simplification of training data to new, unseen data, training datasets incorporate a variation that carries data and variations that are irrelevant to the problem at hand (Duboue, 2017). Feature engineering may also be a method that helps to get options that are general and versatile, so that they may be reused in alternative models for alternative functions. Trained systems are difficult to make and memorizing the proper options for coaching the system is the most crucial and time overwhelming part of developing a productive classification or predictive model (Rawat, 2017).

Moreover, feature engineering is completed when data cleanup and preparation before coaching the model. The main purpose is to supply a better illustration of the dataset to the prophetical learning algorithm rule. Hence, engineering an honest feature set may be a requirement to reach high accuracy in classifying objects and predictions (Rawat, 2017). Data preprocessing is one of all the various crucial steps of any Machine Learning project. As it is known, the real-life information is usually terribly unorganized and noisy and while not data preprocessing, there is no that means of creating a machine learning model (Goyal, 2021). Some of the preprocessing methods are data exploration to identify the data, exploratory data analysis, which include obtaining aware of the variability, distribution, statistical performance of the various input feature, presumably several of their interaction and their relation to the target class or worth we tend to are attempting to predict or regress. Missing data imputation and handling outlier, feature transformation, and so on. Nevertheless, the feature engineering process proposed here are feature transformation and features creation. As the part preprocessing is done in part 2, which includes feature transformation. Now, the dataset is clean.

# **Feature Transformation**

Feature engineering is the task of rising prophetical modeling performance on a dataset by transforming its feature area (Nargesian et al., 2017). The modification created within the format, or the structure of the dataset is named data transformation, this step will be straightforward or complicated supported by the necessities. In other words, data transformation could be a technique that is usually used despite the model it accustomed clarify, whether it is a classification problem or regression problem, or be it an unsupervised learning model. Logarithm transformation (log transformation) is one of the foremost unremarkably used mathematical transformations in feature engineering. It assists in handling skewed data and after transformation; the distribution turns out to be a lot of approximate to normal. It additionally decreases the result of an outlier, because of the standardizations of magnitude variation and the model becomes additionally strong. In this study, feature transformation is done in part 2 using a natural log and standardized data. The result shows that the data is normally distributed, and the outlier is decreased. However, in this part recode the target will be involved.

## Recode Range

In this part, we recoded the target variable “Income” into income >=50000, <=100000, and >100000, the result demonstrates that most the American earn revenues inferior to 50000$, but fewer of them earn an income superior to 10000$.

Graphical user interface, text, application, email

Description automatically generatedChart, bar chart

Description automatically generated

# **Feature Creation**

Feature creation is one part of feature engineering, it is a method of creating a new feature from the present dataset to train a machine-learning algorithm because usually, the raw dataset gathered from different sources will not have the desired variable to perform machine learning task. Feature creation is that part of machine learning that’s thought about further as an art than a science as a result it involves individual involvement in innovatively mixing the present feature (Mueller & Massaron, 2016). The well-created new feature will generally capture the vital info in an exceedingly data much more efficiently than the initial feature. In this study, the feature is not employed.

# **Conclusion**

To conclude, feature engineering is a method used to prepare data for machine learning. it is used once the data is clean, that means after preprocessing. Therefore, is composed of two different methods of feature transformation and feature creation. as the preprocessing is done on part 2, his paper just proceeds with a recording of income variables demonstrating that most American earn an income <=500000.

# **REFERENCES**

Duboue, P. (2017). *Feature Engineering Human-in-the-Loop Machine Learning*. 109–110.

Goyal, C. (2021, October 26). *Feature Transformations in Data Science: A Detailed Walkthrough*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/05/feature-transformations-in-data-science-a-detailed-walkthrough/

Mueller, J. P., & Massaron, L. (2016, October 6). *Machine Learning Creating Your Own Features in Data*. Dummies. https://www.dummies.com/article/technology/information-technology/ai/machine-learning/machine-learning-creating-features-data-226788/

Nargesian, F., Samulowitz, H., Khurana, U., Khalil, E. B., & Turaga, D. (2017). Learning feature engineering for classification. *IJCAI International Joint Conference on Artificial Intelligence*, *0*(August), 2529–2535. https://doi.org/10.24963/ijcai.2017/352

Rawat, T. (2017). Feature Engineering (FE) Tools and Techniques for Better Classification Performance. *International Journal of Innovations in Engineering and Technology*, *8*(2). https://doi.org/10.21172/ijiet.82.024