Correlation Analysis

Correlation analysis determines the strength between two variables, which are commonly referred to as dependent and independent variable. It is a technique used to filter out variables that do not have a strong or significant relationship with the target variable. The strength of the relationship is defined by a decimal variable. The range of the value is between -1 to +1. Correlation Analysis is a great way to determine the strength and direction of the relationship between the two variables. Higher number of coefficient means that the two variables are positively correlated whereas lower number of coefficient means that the two variables are negatively correlated.

A common hypothesis we have for our project is that, attributes that rank higher in correlation coefficient will have stronger ability in predicting classification accuracy. Therefore, we use correlation coefficient to rank every attribute and filter out correlation that do not meet the threshold. This helps to reduce a lot of processing power and saves us a lot of time before training our model. The limitation of this analysis is that in the scenario of two independent variables, it cannot determine the causality between the two. In this research, we only identify the correlation between the utaut factors and moderator factors with our target variable to determine strong predictor attributes.

**Pearson Correlation Coefficient**

The product movement correlation was first proposed by Francis Galton in 1880s, then later modified by Karl Pearson in 1896, and has since been known as the Pearson product-movement correlation. It is a measure that assesses the association between two continuous (or metrical) variables (Statistic Solutions, n.d.).

Below is the formula for Pearson’s Correlation

Diagram

Description automatically generated

Figure 4.2.1 Pearson's Correlation Formula

The correlation values for Pearson’s correlation is between -1 and 1. The closer the value to -1 or 1, the stronger the association. Positive values represents positive relationship, whereas negative values represents negative relationship.

Since Pearson’s Correlation determine linear relationship, therefore linear form is considered to be analyzed among the variables. For correlation among UTAUT Factors, the independent attributes come from itself and moderated Variables. In this case, the relationship is identified based on direction, form, strength of the dispersion of data points between the two parameters. However, pearson’s correlation is commonly used for continuous variables. As UTAUT Factors are scaled variables, the result from Pearson’s correlation coefficient is not as significant. Nevertheless, we can try and compare the results with other correlation coefficient.

**Spearman Correlation Coefficient**

In [statistics](https://en.wikipedia.org/wiki/Statistics), **Spearman's rank correlation coefficient** or **Spearman's *ρ***, named after and often denoted by the Greek letter (rho) or as r s {\displaystyle r\_{s}} , is a [nonparametric](https://en.wikipedia.org/wiki/Nonparametric_statistics) measure of [rank correlation](https://en.wikipedia.org/wiki/Rank_correlation) ([statistical dependence](https://en.wikipedia.org/wiki/Correlation_and_dependence) between the [rankings](https://en.wikipedia.org/wiki/Ranking) of two [variables](https://en.wikipedia.org/wiki/Variable_(mathematics)#Applied_statistics)). It assesses how well the relationship between two variables can be described using a [monotonic function](https://en.wikipedia.org/wiki/Monotonic_function).

Other than Pearson’s Correlation, Spearman Correlation is encouraged to be used for Likert-Scale Response item. This is because Likert-Scale are considered to be ordinal factors, therefore, we should perform non-parametric measure to compute the correlation of the likert-scale factor items, in which case, the UTAUT items.

Spearman’s q (a special case of Pearson’s r)is a nonparametric measure that outputs the correlation using the ranked scores of the two variables (Statistic Solutions, n.d.). It is commonly used for ordinal variables.

Spearman’s q s calculated based on this equation:

A picture containing text, clock, watch, gauge

Description automatically generated

Figure 4.2.3 Formula for Spearman Correlation

The output of the values of Spearman Correlation lies between -1 and 1. Spearman’s q can be interpreted as the difference of normality and proportion of validity between the two ordinal variables.

**Results**

**Correlation network Graph**