# AIML Feature ENGINEERING Projet 1

Weightage: 30%

Indian industries are giving direct Rising inclination towards requesting food online mostly because of expanding digitization, developing utilization of PDAs, portable applications, tablets across the globe, high accessibility of web across a few spots in a locale just as developing revenue towards prepared to eat food are the main considerations that are pushing the development of global Online Food Delivery and Takeaway market over the examination time span.

Further, rising accommodation and less season of delivery brought about by the food delivery organizations are likewise adding to the development of global Online Food Delivery and Takeaway market in the coming years. Moreover, growing convenience caused by the online methods of payments as well as increasing number of discounts and offers provided by the major online delivery companies are also fueling the expansion of the business space.

The best food delivery services ensure that whatever you're in the mood for — pizza, sushi, or something from your favorite local restaurant — it will wind up at your front door without you ever having to venture outside. All it takes is a few clicks on your computer or taps on your phone to get breakfast in bed, a healthy lunch, a spread for the whole family or a late night snack.

Food delivery services have become very popular in recent years, but never more so than during the past 12 months when the coronavirus pandemic has made dining out a no-go for many. Even with restrictions easing up on dining in some areas, food delivery apps are in no danger of being deleted from our smartphones any time soon.

Thats why many food companies are venturing out into this space and representing themselves as competition to the existing players in this segment. One of the key to succedd in this space will be better customer satisfaction rate, which can prove as vital factor while maintaining the existing customer base.

"SwiggyTo" is one of such existing player which has identified this in early stage of the competition and hired you as a data analyst who can help them in identifying the important trends going on with their existing customer base. For that purpose, they have made a small portion of their order database available to you, which you are expected to analyze thoroughly and come with the hidden insights which can provide specific direction to the management in order to imrpvoise their customer service further.

As a part of this project, you will be making use of datasets provided and help identifyign the hidden insights from it. Also you are supoose to identify the critical factors that are impacting the customer service. The main objective of this project is to give you real life experience while doing data acquistion, data integration, data cleaning and data transforamtion before attempting any of the analytical activity on the data.

the various tasks that you will be doing as a part of this exercise will be as follows:

1. Data Qualityand Transformation

beafore the model is built, it is essential to ensure the quality of the data for issues such as reliability, completeness, usefullness, accuracy, missing data and outliers.

data imputation techniques may be used to deal with missing data. use of descriptive stat and visuzlaization may be used to identify the existence of outliers and variability in the dataset. Many new variables can be derived and alos used in model building. categorical data has to be pre-processed using dummy VARIABELS, before it is used to model building.

* you have to cleanse your datasets to remove all such daunting isssues.
* narrate all the issues which you encounter during this exercise clearly with appropriate explanation and code.
* the weigtage for this task will be 17 marks.

1. perform descriptive analytics on data

it is always good to perform descriptive analytics before moving to building a predictive analytics model. it will help to understand the variability in the model. it also helps in identifying the relationship between the variables present in the dataset.

* use the exploratory data analysis technique on the dataset in order to find out the interesting insights that are hidden within the data captured.
* describe all eda steps those are done with the observations obtained out of it with the help of python code in jupyter notebook.
* the weigtage for this task will be 6 marks.

1. feature engineering

Feature Engineering is an important step to develop and improve performance of Machine Learning models. these techniques can also be used to identify the variables that impacts the outcome of the model. its basically Process of identifying and extracting the useful features from the available data. the Primary goal is to derive a set of features that best represent the insights hidden in the data, with a simpler model that generalizes well to future (unknown) observations.

* in this step you have to use knowledge of feature selection methods to identify the variables that have greater impact on the outcome.
* ellaborate description of the impact observed is expected as outcome of this step.

1. model buildingand diagnostics

in this stage first data is divided into train and test data. the subset can also be created using random / stratified sampling procedure. this is important step to measeure the performance of model using dataset not used in model building. it is also essential to check for any overfitting of the model. then model is built using training dataset to estimate the model parameters.

* you have to prepeare a regression model to predict the rating of a order..
* find the significant features from the above model and build another regression model with only the significant variables.
* compare the performace of both models using various model attributes and recommend a model that can be used by “SwiggyTo”
* the weigtage for task C AND D together will be 7 marks.

Notes:

* This is a take-home PRoject to be carried out by group OF LEARNER.
* as per the need, the demos / vivas can be arranged further on.
* wherever required you can make appropriate assumption but make sure that you have spelt them appropriately in the submitted documents.
* This is programming exercise - requiring the approach of appropriate model BUILDING.
* You may consult / discuss with other learners peripheral aspects such as the environment but not on solving the specific problems in terms of design or implementation.
* You have to write the appropriate Python code in Jupyter notebook to support youR answers and submit with following nomenclature

- FE\_project1\_<BITS\_ID>.ipynb

* In case of any further queries, if those are generic once, learners are encouraged to use discussion forums, otherwise they can reach out to US at [amitkp@wilp.bits-pilani.ac.in](mailto:amitkp@wilp.bits-pilani.ac.in), [radhika.richy@wilp.bits-pilani.ac.in](mailto:radhika.richy@wilp.bits-pilani.ac.in), [sujatha.s@wilp.bits-pilani.ac.in](mailto:sujatha.s@wilp.bits-pilani.ac.in) .
* Manage your efforts properly as there is no scope to shift the deadlines announced above.

References:

1. Linear regression concepts learning in course 1
2. Lessons learnt in course 2