|  |  |
| --- | --- |
|  | |
| ECommerce Purchase Prediction  Using  Machine Learning | |
| 23rd April 2024CIS 550 |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | OverviewThis project is related to CIS-550 Advanced Machine Learning course. The present report starts with a general idea of the project and by representing its objectives. Then the given dataset will be prepared and set up.An exploratory data analysis is carried out in order to develop a machine learning algorithm that could predict whether an e-commerce transaction has taken place or not until a final model. Several Algorithms will be used to compare which dataset is more compatible with our dataset and then techniques like smote and Nr can be used to perform over sampling and under sampling on the data respectivelyResults will be explained. | |  |
|  | Project Proposal | |  |
|  | The project intends to train machine learning models to predict whether a customer will buy a certain type of product listed on an eCommerce website or not. The models will help websites designer to create better sales strategies i.e. offering discounts and offers to customers while they are browsing for products in real-time.  To do this, we will rely on session and visitor information and use the various machine learning techniques to reach the target audience, offering them the most relevant products for purchase. In this project, we have used 6 different models: Naive Bayes, Logistic Regression, | Random Forest, Decision Tree, Support Vector Machine (SVM), and k-nearest neighbors. We have also divided our data into various testing experiments to observe the changes upon standardizing, under sampling for majority class oversampling for minority class, and computed their results later in a confusion matrix to obtain various quantities    such as accuracy, success rate, sensitivity, and other such parameters of the model finding out the best model for each experiment. The techniques and coding part have been done in Python and executed using a Jupyter notebook. |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | Domain Knowledge: Technology has enabled people to connect more quickly, effectively, and reliably. With each mass implementation of technology to people, each implementation has made greater strides into making life more comfortable and convenient for them. Ecommerce has been one such congregation of technologies and processes which allow commerce to happen in an online form.  The online form has allowed for faster transactions, greater selection, and better customer support. All these conveniences have been beneficial for the customers at large. With the COVID-19 pandemic, it has become clear that such a form of commerce has become the norm rather than the second counterpart. | The COVID-19 pandemic resulted in an increased demand for doing business via the Internet. This has resulted in a search for tools and techniques for better prediction of sales and new marketing strategies by managers of retail firms, both online and brick-mortar going online. Now, they want to orient their marketing actions to the right target. Now a newer trend has emerged of identifying in real-time the purchasing intention of potential customers as soon as they have logged onto the website.  Most want insights from customer behavioral data of online customers which help them formulate these strategies. For acquiring these insights they are turning to experts in data analysis, data science, and machine learning fields. |  |
|  | Image | |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Literature Review  Literature encompasses many studies which are turned towards categorization of online visits in  e-commerce websites.  A first study of Mobasher et al. [2] assessed two different clustering techniques based on user transactions and pageviews in order to find out useful aggregate profiles that can be used by recommendation systems to achieve effective personalization at early stages of user’s visits in an online store.  Later, the study of Moe [3] prepared the ground for a system which can take customized actions according to the category of a visit. For this reason, the author proposed a system which makes use of page-to-page clickstream data from a given online store in order to categorize visits as a buying, browsing, searching, or knowledge-building visit. The proposed system rests on observed in-store navigational patterns (including the general content of the pages viewed) and a k-means algorithm for clustering.  In [4], Poggi et al. proposed a system which handles the loss of throughput in Web servers due to overloading, by assigning priorities to sessions on an e-commerce website according to the revenue that will generate.  Data were formed of clickstream and session information and Markov chains, logistic linear regression, decision trees and naïve Bayes were investigated in order to measure the probability of users’ purchasing intention [4].  In [5] and [6], authors designed the prediction of purchasing intention problem as a supervised learning problem and historical data collected from an online bookstore were used to categorize the user sessions as browsing and buyer sessions. |  |  |