Intro to AI Report

Project Members: Harpal, Deyan, Mahmood

• Report, introduction: description and motivation of the problem, description of the dataset including data types (e.g. discrete, continuous) (15%)

What is your dataset, problem domain?

• Did you have any missing data? If so, how did you cope it?

• Did you do apply techniques to understand your dataset?

• Have you omitted some data? If so, why?

# Project Overview

Many online retailers continually try to create solutions in order to increase the traffic of users that end up purchasing an item through their website. This could be through many different techniques such as the aesthetics or usability of the website itself. Machine learning may also help with this outcome by finding correlations between the statistics of website use, and using this information, be able to create a sustainable plan to increase revenue.

# Problem Domain

The goal of this project is to use different machine learning algorithms to find a way to increase the revenue of online retailers. The tasks involve:

1. Downloading the Online shoppers’ intentions dataset
2. Using Data visualisation to identify the links in each data categories
3. Apply Machine learning algorithms to this data
4. Outputting the results of this data
5. A conclusion on the results

# Analysis

## Data Exploration

The Online shoppers’ intentions dataset has a wide dimensionality of 18 columns and contains over 12000 rows of user data. The dataset contains 10 numerical and 8 categorial attributes respectively. The ‘revenue’ column will be used as the class name, as we are trying to find out how to increase the revenue. 7gYTGKCJvB33nJx

A close up of text on a black background

Description automatically generated

Figure Columns and their data types, taken from project code.

## Fields from Online shopper’s intention dataset.

* **Administrative, Administrative Duration, Informational, Informational Duration, Product Related** and **Product Related Duration** represent the number of different types of pages visited by the visitor in that session and total time spent in each of these page categories.

The **Bounce Rate, Exit Rate** and **Page Value**  represent the metrics measured by "Google Analytics" for each page in the e-commerce site.

* **Bounce Rate**- feature for a web page refers to the percentage of visitors who enter the site from that page and then leave ("bounce") without triggering any other requests to the analytics server during that session.
* **Exit Rate**- feature for a specific web page is calculated as for all pageviews to the page, the percentage that were the last in the session.
* **Page Value**- feature represents the average value for a web page that a user visited before completing an e-commerce transaction.
* **Special Day**- feature indicates the closeness of the site visiting time to a specific special day (e.g. Mother’s Day, Valentine's Day).The value of this attribute is determined by considering the dynamics of e-commerce such as the duration between the order date and delivery date. For example, for Valentina’s day, this value takes a nonzero value between February 2 and February 12, zero before and after this date unless it is close to another special day, and its maximum value of 1 on February 8.
* The dataset also includes operating system, browser, region, traffic type, visitor type as returning or new visitor, a Boolean value indicating whether the date of the visit is weekend, and month of the year.

( This has been taken from the dataset description)

• Report, methodology: summary of the models used, with their pros and cons, a hypothesis statement, description of choice of training and evaluation methodology (20%)

• What models did you use?

• Is your model classification or regression?

• Report, results: description and presentation of the output. The code acts as an appendix to this section, and code quality (e.g. commenting) contributes. (35%)

• How did you encode the input variables?

• What are the criteria for selecting model performance evaluation tools?

• What were your outputs?

• Report, evaluation: analysis and critical evaluation of results. (10%)

• Did you have any problems or difficulties working with the dataset?

• Report, conclusions and referencing: lessons learned, references (using Harvard format) and future work. (10%)

REFERENCES PAGE