**Professional Certificate in Machine Learning and Artificial Intelligence**

Module 25: Portfoglio Project

Datasheet for Dataset Template

**Motivation**

Answer the following questions:

* For what purpose was the dataset created?
  + Answer: This data studies whether a person will accept the coupon recommended to him in different driving scenarios for marketing purpose
* Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)? Who funded the creation of the dataset?
  + Answer: Data was collected by University of IOWA researchers.

**Composition**

Answer the following questions:

* What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?
  + Answer: Data set is composed of categorical data and integers. Dataset is about marketing coupons for truck drivers through mobile marketing.
* How many instances of each type are in total?
  + Answer: 12684 instances
* Is there any missing data?
  + Answer: Yes, Car column is 99% missing data and there are other 5 columns with the missing data
* Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor–patient confidentiality, data that includes the content of individuals’ non-public communications)?
  + Answer: No data set contains survey data on Amazon truck drivers. There is no specific reference to the personal data in this survey

**Collection process**

Answer the following questions:

* How was the data acquired?
  + Answer: Data was acquired through survey on truck drivers
* If the data is a sample of a larger subset, what was the sampling strategy?
  + Answer: This is complete data which has been cleaned. Truck drivers with > 95% rating were selected for this survey. Out of 752 surveys 652 were accepted
* Over what time frame was the data collected?
  + Answer: There is no reference to the time line of the data, dataset was first made public in 2015

**Preprocessing/cleaning/labelling**

Answer the following questions:

* Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.
  + Answer: Missing rows were eliminated other than that no preprocessing was done
* Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)?
  + Answer: No

**Uses**

* What other tasks could the dataset be used for?
  + Answer: Dataset can only be used for coupan ecommendation system on the specific routes for Amazon truck drivers
* Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other risks or harms (e.g., legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?
  + Answer: Dataset should only be used for mobile recommendation model on specific routes of Amazon truck drivers
* Are there tasks for which the dataset should not be used? If so, please provide a description.
  + Answer: it should only be used for mobile marketing

**Distribution**

Answer the following questions:

* How has the dataset already been distributed?
  + Answer: Data set has been distributed under CC 4.0 license and is available to public since 2015
* Is it subject to any copyright or other intellectual property (IP) license, and/or under applicable terms of use ToU)?
  + Answer: The license allows for the sharing and adaptation of the datasets for mobile marketing purpose, provided that the appropriate credit is given.

**Maintenance**

* Who maintains the dataset?
  + Answer: At the moment dataset is not maintained.