Report of interview task

Wish list project

Fatemeh Fazeli Asl

November 2022

Abstract

Wish list project looks forward rating the vendors of Gorgan city in terms of quality level of services. Same as all project, it needs to move across some practical steps to end up with valid results. Firstly, all critical information of vendors will be collected from available sources so that it could be possible to make a good database of vendors. For example, web search engines can help to find pure data of location, phone number, name, etc. of the vendors, and they also introduce another useful sources to gain related data. Secondly, it is necessary to build a clear database of all gathered information in which it clarify data by separated columns. Finally, the database is ready to create classifier models for grouping the vendors based on the target of quality level. A brief description of this project has been presented in the following.

1. Source of vendors' information

There are several sources which can give us a hand in the case of collecting a great database about top vendors in Gorgan city of Golestan. Some of the most effective sources and a brief explanation of how they work, are listed in the bellow.

- **Web search engines:** The first and immediate way to obtain any kinds of information all over the world is to search the related keywords through web engines like google. Therefore, we can utilize this method to find suitable information about vendors such as location, type of services, products, name, capacity, phone number, service's price, web-page (if existed), etc. I use this source to obtain some information about food vendors of Gorgan city.
- **Google map:** To find out how much vendors of a certain good or service like restaurants are existed in an overall scope with geographical details, google map or other online map can be the best tool. It is worth to mention that since the online maps are mostly updating by free users or owners of the recorded geographical points, they might not be valid or completely true. In some cases, we cannot find real vendor's location entirely on the map because they might have not registered by vendors, before. To do this task, I gathered a list of food vendors of Gorgan city in different types of fast food, classic and modern restaurant, etc. through google map in which it consists of valuable information like location, name, phone number, services options, etc.
- **National and local dependent organizations:** There are many organizations that are responsible for monitoring the population of vendors based on their products or services. Most of the time, these organizations propose periodical reports of vendors performance with details in their websites or in the form of published magazines. For example, we can find a list of food vendors of Gorgan in the website of https://iranianasnaf.ir/ which is assigned to Organization of Iranian Asnaf.
- **Social media platforms:** Nowadays, we can find a vast variety of Instagram and similar social media pages which announce business advertisement of all kinds of products' or services' providers. A range of food-blogger Instagram pages could be a well-worth source for collecting information of Gorgan's food vendors. In this report I couldn't use this source because of restricted access to Instagram in political reasons.
- **Live observations:** To validate and complete all gained information from all another sources a trustworthy way is to visit and walk around vendors' location and fill manually all blanked or missed data in constructed vendors' database. Since I hadn't enough time and facility, I couldn’t do such a live observation for doing this task and I only relied on received information from google map, web search engines, and related websites.

- **Questionnaires:** A great and necessary way to know top vendors in each city is to fill some suitable questionnaires by people who live over there or are consuming the vendor's products. This way needs a series of functions which cost a long time, therefore I couldn't utilize this source at all.
- **Advertising centers or distributed advertisements:** online, or offline published announcements are another potential source to know what kinds of vendors have being shown-off their services or products. Only by taking a look at the history of vendor's advertisements we can obtain valuable information about its evolution or developments.

To a nutshell, I constructed a dataset of top food vendors of Gorgan city by searching in google map and related websites. Since google map provided a rating system to each location based on user comments, it would be possible to choose top rated vendors by selecting vendors with 4 or more than 4 stars.

2. Vendors' database

The mentioned dataset in the previous section consists of name, type of service (e.g. restaurant, fast food, etc.), location, phone number, closeness degree to the center, rating, service option, type of served meal, opening and closing time, work time, the number of users who rate the vendor. Some of these features are explained in the bellow.

I divided Gorgan into three region to certify the closeness degree of each food vendor to the center of Gorgan. The main reason of this idea is to convert location data to a certain numerical value so that it could be used as a feature to train the classifier model. As it can be seen in Figure 1, the entire shape of Gorgan has divided into three region by three different boundary lines. Third region is displayed by green line which shows the center of Gorgan, second region is the blue one as the middle area of Gorgan, and the first region has been displayed with red line which is called margin area. Since usually the center region of cities is the most popular place to visit, it has been assigned by the value of '3'. The last two region are assigned by the value of 2, and 1, respectively because they worth less than the center region.

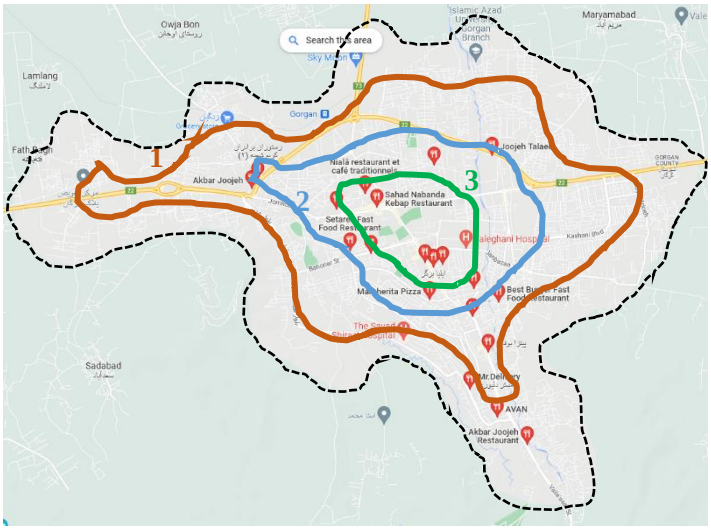

Figure 1: Divided Gorgan into 3 region based on closeness degree to the center

In the case of service options, there are three kinds of options that vendors provide for customers namely: dine-in, takeout, and delivery. Some of vendors use all the mentioned options and the others use one or two of them. If we suppose the numerical value of dine-in, takeout, and delivery is 1, 2, and 3, then the

combination of them like dine-in & takeout, dine-in & delivery, takeout & delivery will convert to 12, 13, and 23 numbers, respectively. Finally, all these options i.e. dine-in & takeout & delivery will take them a value of 123. Such this method has been implemented for assign numerical value to 'Meal' column.

The column of work tine has been built by calculating the differences between opening and closing times of vendors. This feature is important from the aspect of service duration to customers. Whatever a food vendor serves in a widest duration of the day, customers can have their meals every time they want.

It is necessary to mention that the rating of the vendors was done by users who voted to each vendor within google map's website. Since some of vendors have been voted by only one or a few number of users then they saw a high rating. Therefore, the value of rating per each vendor is not as trustworthy as we may think. Consequently, the number of reviewers comment can be a validation base of vendor's rating.

3. Model classification

With the vendors' database in hand, we can classify all vendors into different classes of quality. First of all, we need to organize and clean the database so that it can be used to learn the classifier model effectively. In this project the vendors' database has been constructed from 20 rows and 12 columns. In Table 1 a part of this database is shown.

Table 1: A cut part of vendors' database

Name	Type	Location	Closeness degree to the center	Phone number	rating	Service options	Meal	Opening time	Closing time	Work time	reviewers comments
Melal	Restaurant	RCMH+XMW, Gorgan, Golestan Province, Iran	3	1732247400	4.1	Dine-in & Takeout & Delivery	full	8	24	16	162
Bazaar	Restaurant	Golestan Province, Gorgan, Shahid Rajaei St, RCRF+J3H, Iran	2	1732430300	4.4	Dine-in & Takeout & Delivery	Lunch & Dinner	8	24	16	182
Diners	Restaurant	VC7X+462, Gorgan, Golestan Province, Iran	1		5			8	18	10	1
Nialâ restaurant et café traditionnels	Restaurant	RCVH+H63, Gorgan, Iran	3	9394591817	4	Dine-in & Takeout	Lunch & Dinner	8	24	16	2
Arcadas Turkish	Restaurant	Gorgan,2nd Homayon، RCMR+9X2, Iran	3	6117012387	4	Dine-in & Takeout	Lunch & Dinner				1
Pialeh	Restaurant	Golestan Province, Gorgan, 11th Edalat ، RCMQ+FM7, Iran	3	1732345874	4.5	Dine-in & Takeout	full	8:30	23	14.5	10
Setareh fast food	Fast food	RCPG+44V, Gorgan, Golestan Province, Iran	3	1732226383	5	Dine-in & Takeout	Lunch & Dinner		24		2
Joojeh Talaee	Restaurant	Golestan Province, Gorgan ،Easet Jahad, RFX2+QMR, Iran	1	1732154889	4.3	Dine-in & Takeout & Delivery	Lunch & Dinner	9	23	14	78

Hence, we start the pre-processing procedures with removing all useless columns like name, phone number, location, type, opening and closing time. It is worth to mention that location and opening/ closing times have been converted to a numerical form because they have a slight effect on the classification model. All nun-numerical data must be converted to a numerical format in our database including: Service options, Meal. Moreover, all the missing values in all columns must be filled with valid data. There are some missing values in the column of work time that we must find a way to fill them. Since the number and type of meal effect on the work time of vendors, new values of missed work times could gain by averaging the amount of existed work time per each meal type.

The pre-processing process has been just finished but we need to define a target label so that we can classify data based on target label. The best target for this database is rating column but we convert it into a new form in which the vendors with more than 4.5 rating labeled as level '1' and the others with '2'.

Finally, we can split the database to features including: closeness degree to the center, Service options, meal, work time, reviewers comments, and target namely rating. In this example, 70 % of data are used as train data and 30% of them are used as test data. The classification model of Naïve Bayes is selected as the first suitable model for this database because of the small volume of data. As well as, logistic regression model is used for training data to compare with the first model. According to Table 2, the model accuracy of logistic regression is greater than Naïve Base in this example. Therefore, it is better to utilize the logistic regression model for classifying all vendors to different level of quality.

Table 2: Comparison of classification model accuracies

Model	Accuracy
Naïve Bayes	61.7%
Logistic regression	73.3%

4. Conclusion

Starting with gathering related information from all available sources to end up by classifying data based on a certain target, a small data analyzing project namely wish list project is done in this report. The main aim of this project is to certify Gorgan's vendors by two level of quality 1, and 2. To do this, two famous classification model namely logistic regression and Naïve Bayes is used to classify the vendors' database to quality level of 1 and 2. By comparing gained results, logistic regression is selected as the more accurate classification model in this project.

References

https://github.com/fatemefazeli/Snappfood.git