Disease Outbreak Prediction Using Data Mining and Machine Learning

Alaukika Diwanji, Dhruwaksh Dave, Panth Desai and Udit Marolia

***Abstract*—Many valuable information regarding the public health and welfare, disease outbreaks and their trend are available in the form of unstructured data lying in different news portals, Facebook, Twitter. It becomes important to become aware of the current diseases and to filter out relevant and correct information. This is especially important for commercial pharmacies as their need to be updated with the current outbreak in their region and also be ready stock-wise for the drugs needed to treat them. Our objective with Well-Pharma is to address this problem and built a system for the pharmacies which will analyze the disease outbreaks in all regions and carry out a disease-to-drug mapping and alert the pharmacist so as to keep the stock ready.**

WellPharma will be a Web Application - built as an automated system for querying filtering and visualizing the disease outbreak and to stock their respective drugs.

INTRODUCTION

Today, one of the major challenges faced by the people is to seek information on Health. We need some useful information related to this department that can be helpful to solve some major issues like the one of the pharmacies that we have introduced in this paper. It is best that we solve this using the most advance and emerging technology, Machine Learning. The problem concerned is here of the pharmacies that do not have any idea of which diseases may prevail for the following period and they are not ready for those. The price of the drugs increases in such outbreak\s and the customers as well as the pharmacies have to suffer a lot because of that reason. This problem can be overcome only if the pharmacies are aware of what diseases will be occurring in the neighborhood in near future. Our main goal will be to provide the pharmacy with the accurate predicted values of the diseases.

MATERIALS AND METHODS

The solution needed was to create a prediction system for the pharmacy that will provide them with the forecast of which disease will prevail for the next month.

This was done using data mining and machine learning. Firstly, a crawler was created that crawled through the CDC website to fetch all the name of the diseases. We got about seven hundred diseases which contained some noise in it. This had to be solved manually. Resulting this was a list of 237 diseases. After this, we created a database that contained the drug list for found diseases. Another crawler was created which surfed through the internet to find the "curing drugs" for the diseases. The crawler was able to find some drugs for the diseases. Later the other ones were manually searched and added to the database. This database is on Mongo Atlas. On the model part, we used the Google Trends API to get the data of the number of searches according to the DMAs on the monthly basis from the year 2010-19. The data was stored in the comma separated values type file. This data was sorted to get top 5 diseases searched from all the diseases. From this, we got 31 top diseases. The search was made again to get on the 31 sorted disease using the same Google Trends API to get the monthly number of searches from the year 2015-19.

This data of the 31 diseases was then send to the prediction algorithm using the Facebook Prophet. The resulted outcome was the data of the predicted values from between 0-100 which were stored in the data base along with the diseases and the mapped drugs. The higher the values, more the chances of those diseases to prevail in the coming month. These values are sorted based on the descending values of the predicted values according to each DMA. This all data is stored on the mongo. This all can be easily fetched using the mongo queries and later, on the UI when the user searches for a particular DMA, the query would run in the backend and this would provide the search results in the table format to the user.

This is also shown in the map format without the information about the drugs to the normal user who isnt registered.

THE PROPHET FORECASTING MODEL

There is a wide diversity of business forecasting problems, however there are some features common to many of them. Importantly, it is also designed to have intuitive parameters that can be adjusted without knowing the details of the underlying model. At its core, the Prophet procedure is an additive regression model with four main components:

1.A piecewise linear or logistic growth curve trend. Prophet automatically detects changes in trends by selecting change points from the data.
2.A yearly seasonal component modeled using Fourier series.
3.A weekly seasonal component using dummy variables.
4.A user-provided list of important holidays.

The important idea in Prophet is that by doing a better job of fitting the trend component very flexibly, we more accurately model seasonality and the result is a more accurate forecast. We prefer to use a very flexible regression model (somewhat like curve-fitting) instead of a traditional time series model for this task because it gives us more modeling flexibility, makes it easier to fit the model, and handles missing data or outliers more gracefully.

We have used the Prophet model to produce the desired predicted values of each disease. The monthly data of all the

diseases are generated using the trends API and after it, we have trained that data on Prophet.
We use a decomposable time series model (Harvey & Peters 1990) with three main model components: trend, seasonality, and holidays. They are combined in the following equation:

$$y(t) = g(t) + s(t) + h(t) + €t.$$

Here g(t) is the trend function which models non-periodic changes in the value of the time series, s(t) represents periodic changes (e.g., weekly and yearly seasonality), and h(t) represents the effects of holidays which occur on potentially irregular schedules over one or more days. The error term €(t) represents any idiosyncratic changes which are not accommodated by the model; later we will make the parametric assumption that €(t) is normally distributed.

RESULTS

The problem what we solved here is the prediction of diseases that will help users (pharmacy) to get information in advance and store the drugs accordingly. Two kinds of users are can access the web application, normal users and the registered users (pharmacists). The normal users are anyone who wants to view the website. They will be provided with the information on map where they will get the information on which disease might get spread in each DMA of a state selected in the next month. A DMA is a designated marketed area where the population can receive the same (or similar) television and radio station offerings and may also include other types of media including newspapers and Internet content.

The other kind of users are the registered users which for this project are pharmacy owners. The user will be guided to the select region area where he will have to select the DMA, he wants the information about. The output generated will be in the form of a table where there will be top five predicted diseases for that particular DMA and the suggested drugs is shown. According to a study from CBS news, pharmacies have to face losses as the pharmaceutical companies hike the prices of about two hundred and fifty OTC drugs to about six percent. This loss can be majorly prevented using this application.

DISCUSSION

The result that is generated from the prediction model is passed on to the database and then derived t=from the database to the frontend. This data is in tabular format which include the disease in one column and its drugs in the other one. This is very useful data for the persona using this application i.e. the pharmacy. They can view which all diseases may prevail in the next months and also, they have the suggested drugs which they can store in the pharmacy. They can compare their stock based on the requirements and then place order of the required amount in advance to the pharmaceutical companies.

By doing this, the pharmacy as well as the customers will be happy. The pharmacy won't have to spend more money when the price hike occurs and also for customers, they too need not unnecessarily pay more for the needed drugs in the crisis period.

FUTURE ENHANCEMENTS

The WellPharma application is currently serving a single persona i.e. pharmacy owner. In future we intend to increase it to the normal users providing them more than one functionality rather than just showing them the map which will show the disease that might outbreak in the next month in each state.
We also plan to include the pharmaceutical companies as the persona. This will be the most instrumental step as the drugs production can be controlled by them and this will be beneficial for both the normal customers as well as the pharmacy owners as well.

We can also add the functionality for pharmacy owners to directly place orders of the needed drugs to the pharmaceutical companies. We can also predict the beauty trends using the Google Trends API through which the data on beauty can be predicted and cosmetics can be added to the picture.

ACKNOWLEDGMENT

The success and final outcome of this project required a lot of guidance and assistance from many people and we are extremely privileged to have got this all along the completion of my project. All that we have done is only due to such supervision and assistance and we would not forget to thank them.

We respect and thank Mr. Rakesh Ranjan, for providing me an opportunity to do the project work in Enterprise Software Overview on WellPharma and giving us all support and guidance, which made us complete the project duly. The inputs and views provided by him on this project's initial definition helped us create what we have done today. We are extremely thankful to him for providing such a nice support and guidance, although he had busy schedule managing the corporate affairs.

We also thank our friends and seniors for providing us support and help through which we were able to get through each hurdle that came in out way while completing this project. All the suggestions and the guidance we got was very valuable to us and we are very pleased with it.

We are thankful to and fortunate enough to get constant encouragement, support and guidance from all Teaching staffs of Computer Engineering Department which helped us in successfully completing our project work.

CONCLUSION

The WellPharma application is a very useful platform created for the pharmacies kept in mind as the persona. They can easily look for the diseases which might prevail in the coming month and on basis of that they can are also provided with the drugs to deal with them. Using this, the

user can easily eliminate some major losses that occurs during an immediate disease outbreak and because of this, the normal customers can also be happy as they too do not have to spend extra amount for the medicines they need.

REFERENCES

[1] 1.A Review on Predictive Analysis in Data Mining in International Journal of Chaos, Control, Modelling and Simulation (IJCCMS) , 2016

[2] Predictive Analytics in Information Systems Research in *MIS Quarterly*

[3] https://peerj.com/preprints/3190.pdf (Forecasting at Scale)

[4] https://research.fb.com/blog/2017/02/prophet-forecasting-at-scale/

[5] https://www.cdc.gov/diseasesconditions/az/a.html

[6] https://www.drugs.com/medical_conditions.html

[7]. https://www.pharmacytimes.com/contributor/beth-lofgren-pharmd-bcps/2015/