AAI-800 Mid-stage Report

Student: Xiangyang Xu; Advisor: Prof. Shucheng Yu
ECE Department
Stevens Institute of Technology
Hoboken, USA
xxu46@stevens.edu

***Abstract* — This is the mid-stage report of my AAI-800 course, my AAI-800 course is to use machine learning tech to do covid-19 prediction and embed this prediction as a forecast into an android app. At the same time, people who need it can be recommended for influenza vaccination**

Keywords — covid-19, flu, machine learning, deep learning

I. Introduction & Challenges

A. Introduction

From Dec.2019, the covid-19 spread from Wuhan to the world. Millions of people suffered this new virus and ten thousands of people died due to this virus. At the same time, the economy of many regions stopped growing.

Machine learning and deep learning tech are widely used to help people to fight against this kind of new virus. For example, L. Mertz and his team developed an AI-driven tools to quantify Lung Images.[1] At the same time, in people daily life, many apps are developed for keeping people away from Covid-19 virus. J. Berglund and his team developed a software to track the virus.[2]

With the hard work of people and the support of technology, in July and August this year, the epidemic was contained. But now that winter is about to come, the epidemic may recur. Therefore, anticipation and early warning of Covid-19 are necessary. Use an app to help people away from Covid-19 seems a necessary work. Currently, people use machine learning method to predict the trend of the price of goods. Based on the history price, it is feasible to predict future prices. A. Yousefi and his team has used this tech to predict long-term electricity price.[3] Therefore, it is also feasible to use machine learning to predict the future trend of covid-19.

If we can predict the possible outbreak of the epidemic, then forecast it to public. Then we can reduce the number of people infected. This will help the epidemic to be extinguished quickly, which will result in fewer infections and deaths. This can also make the economy recover faster. Same technology can be also used to forecast flu.

B. Challenges

1) Data procession

In the past, I completed the research project by the dataset given by professor. These datasets are all well designed, and I don't need to reprocess these data sets. This time I have to get open-source dataset from the internet.

There're a lot of open-source dataset about covid-19 can be found from Kaggle.com. However, they have different kind of label and features. At the same time, they are not designed for my project. In order to use machine learning method and make the model more accurate. I have to assemble all the datasets so that they become one dataset with a unified label and as many features as possible

2) Improve model accuracy

After that, I need to constantly try various algorithms and models. Then choose a model with higher accuracy. After this, I need to assemble the model. Only in this way, I can get a prediction model that is as accurate as possible.

3) Embed machine learning into an app

Finally, I need to embed the machine learning into an android app. This is also my first try to apply machine learning in a real app.

II. Related Work

A. COVID-19 Future Forecasting Using Supervised Machine Learning Models[4]

Furqan Rustam and his team use the dataset provided by the Center for Systems Science and Engineering, Johns Hopkins University. This dataset include features like latitude, longitude and the number of new inflect people.

They use 4 algorithms, Linear Regression, LASSO, Support Vector Machine and Exponential Smoothing to do future covid-19 forecasting. They set 85% of data as training data and 15% as test data.

ES and LASSO method perform good in forecast new cases with higher score and lower MSE. LR perform not bad and SVM cannot make a good prediction.

B. Trend Prediction of Influenza and the Associated Pneumonia in Taiwan Using Machine Learning[5]

Sing-Ling Jhuo and his team use a different kind of dataset and algorithm to study the trend prediction of Influenza. In their dataset, there are a lot more kinds of features such as temperature, relative humidity, PM2.5 and CO. Then their study dataset has more features.

Then they use a deep learning method, Multilayer Perceptron (MLP), to make model. Their study achieves 77.54% accuracy for the trend of influenza. The accuracy of elder population is especially good. The accuracy get 81.16%.

III. FORMAL DEFINITION OF THE PROBLEM

A. Overall Framework

My research project will contain follow three parts, Data procession, machine learning prediction and forecasting by android app.

Data processing is responsible for integrating data from different data sets. To form a new data set with many features that may affect the prediction results.

Then use machine learning to predict the covid-19 trend in the area. This research project will mainly use supervised learning. Because according to the existing data set, there will not be too many missing features

Finally, the app allows users to enter local conditions. This leads to a prediction based on a machine learning model. Based on this prediction, give a forecast. So as to help people stay away from the virus

The following figure briefly explains how my research project will work.

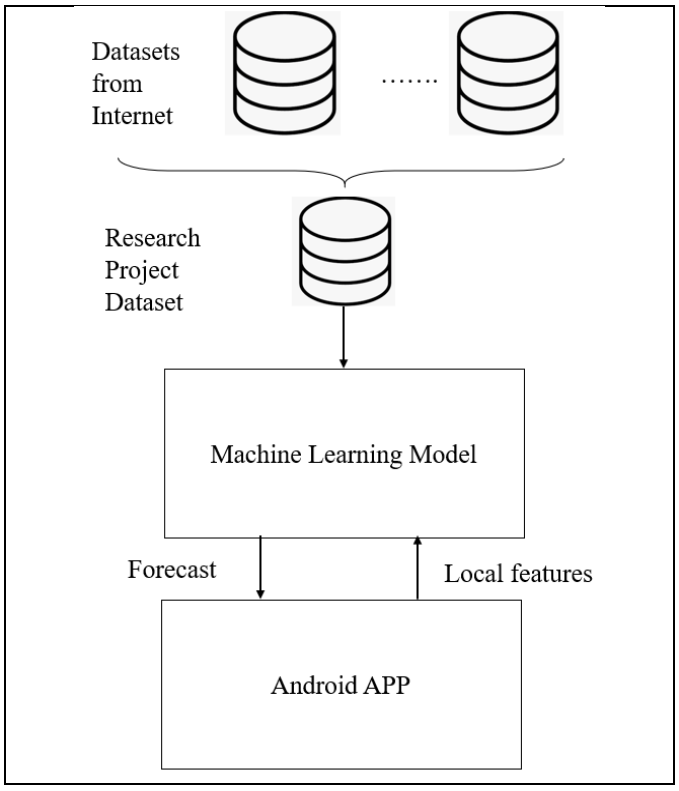

Fig. 1. framework diagram

B. Dataset Procession

Today's datasets on the Internet have different features and labels due to different purposes. Generally speaking, because the purpose of each project is different, the label and feature of their data set are different. It is impossible to build a predictive machine learning model through these original data sets.

This project needs a label and as many features as possible. Therefore, look for the same point in each data set as the label. More features can be integrated, and the machine learning model can make accurate predictions based on this

C. Machine Learning Prediction

A variety of supervised learning algorithms are used to improve the accuracy of each model as much as possible, as well as the error and performance on the MSE.

Then select the best models and assemble them. Then we can get the ideal forecasting model.

At the same time, through the data set of influenza vaccination, we can infer whether the population of this person needs vaccination. So as to provide advice on whether to vaccinate or not.

D. Embed Prediction Model into Andriod APP

In order to make it easier for people to use, in the APP, you only need to select the covid-19 prediction or influenza vaccine recommendation, and you can select the corresponding function.

Then simply fill in some local information, or some personal information. The machine learning model can give corresponding predictions or suggestions.

The user can decide the plan for the day based on the APP recommendations and forecasts.

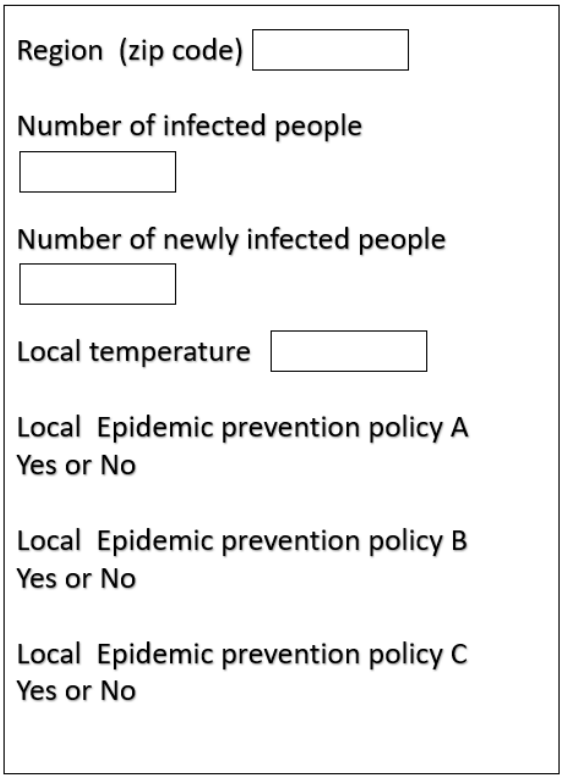

Fig. 2. Sample Covid-19 Prediction Input interface

IV. DESCRIPTION OF THE SULOTION

A. Dataset Procession

1) Covid-19 Dataset

In the data I selected, there is one thing in common, that is, they all have the corresponding address, state, county fips code or zip code.

The state where a county is located can easily find the zip code included. So any feature belonging to the state and zip code can be added to the county.

Then, I can add as many features as possible to the county (fips code) as the label. This way I can integrate as many data sets as I want to use for my research project.

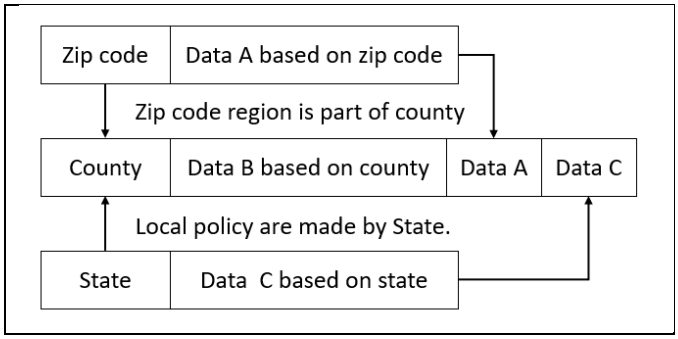

Fig. 3. Data Merge

As Fig.3. shown, I can merge different kinds of data to county label. State will make policy such as "<500 gathering", which will influence county. At the same time, by calculating population, etc., data provided by Zip code, I can get the population of a county.

Then I can get a data set with as many features as possible. I can also predict covid-19 based on county

2) Flu Dataset

There are already mature data sets for influenza health advice, so there is no need to do too much processing

B. Machine Learning Prediction

First, I will randomly divide the data set into training data and test data at a ratio of 70% and 30%. If overfitting occurs, I will divide the training data and test data again.

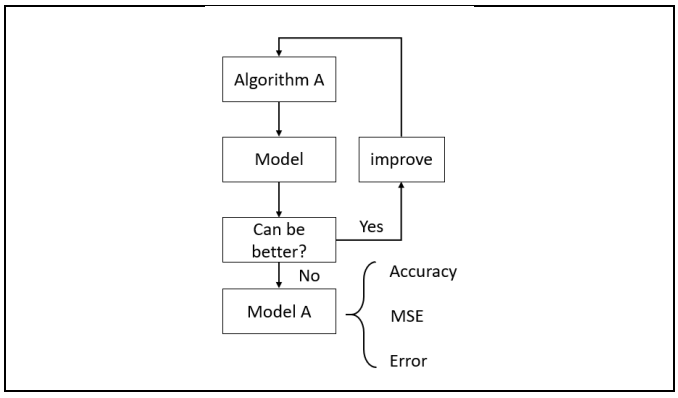

Fig. 4. Model optimization process

This research project will try random forest, linear regression, MLP, clustering and other supervised learning algorithms, and establish a good prediction model as much as possible by adjusting parameters and other methods.

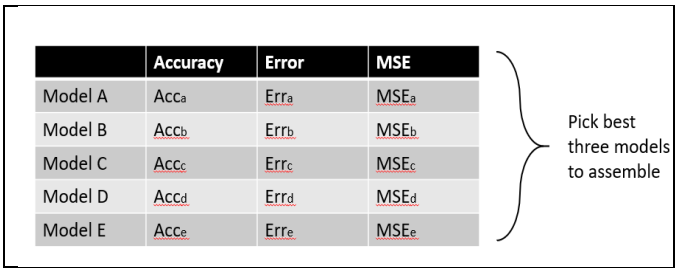

Fig. 5. Select best models to assemble

As is shown in Fig.5., then select several models with the best performance in terms of MSE, accuracy, etc. as the. And use the method of assembly to determine the final model. Thus, this research project can get the ideal expected value.

C. Andriod APP Embed

I will design an app that is easy to use. First, the user needs to choose whether to get flu advice or covid-19 virus advice. Then the user only needs to enter the data provided to the machine learning model to predict and then they can easily get suggestions. This app will be easy to use.

1) Covid - 19 Suggestion

For the convenience of people. The APP only requires people to input related features. For example, the region, the number of infections, the number of new infections, local temperature, local policies, etc. The model can automatically determine the number of new infections that may be added tomorrow and give reasonable suggestions.

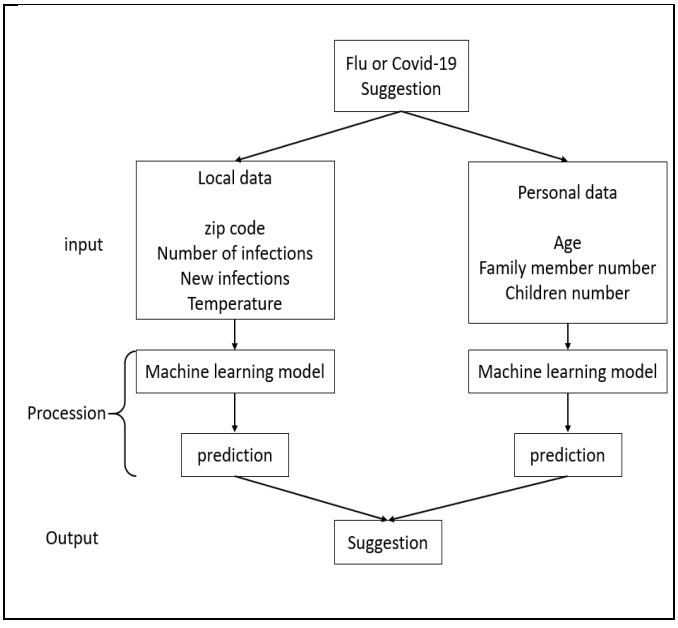

Fig. 6. Suggestion APP Work Diagram

2) Flu Suggestion

Just enter information, such as age, number of family members, number of children, etc. The machine learning model can then give advice on whether to get a flu vaccine.

REFERENCES

[1] L. Mertz, "AI-Driven COVID-19 Tools to Interpret, Quantify Lung Images," in IEEE Pulse, vol. 11, no. 4, pp. 2-7, July-Aug. 2020, doi: 10.1109/MPULS.2020.3008354.

[2] J. Berglund, "Tracking COVID-19: There's an App for That," in IEEE Pulse, vol. 11, no. 4, pp. 14-17, July-Aug. 2020, doi: 10.1109/MPULS.2020.3008356.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] A. Yousefi, O. A. Sianaki and D. Sharafi, "Long-Term Electricity Price Forecast Using Machine Learning Techniques," 2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), Chengdu, China, 2019, pp. 2909-2913, doi: 10.1109/ISGT-Asia.2019.8881604.

[4] F. Rustam et al., "COVID-19 Future Forecasting Using Supervised Machine Learning Models," in IEEE Access, vol. 8, pp. 101489-101499, 2020, doi: 10.1109/ACCESS.2020.2997311.

[5] S. Jhuo, M. -. Hsieh, T. Weng, M. Chen, C. Yang and C. Yeh, "Trend Prediction of Influenza and the Associated Pneumonia in Taiwan Using Machine Learning," 2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Taipei, Taiwan, 2019, pp. 1-2, doi: 10.1109/ISPACS48206.2019.8986244.