**SUICIDE ANALYSIS AND PREDICTION**

**Introduction**

The increasing suicides around the world has become an important concern about studying deep about it and try to understand the reasons behind it.

**Background Problem**

As we it’s been always so heartbreaking listening to each time, we listen to a news about committing suicide.

**Technologies and methods**

Firstly, I want to talk about python dashboard. It’s always wonderful to see how we are able to make models and interpret them. But it is also important to note, recently there are number of concerns about how well we are able to make modifications to the existing model and maintain them. So, our model has to work dynamic and make prediction based on the available data. In recent years programmers used use VueJS or web based languages for making dashboards, we now have most advanced packaged like Streamlit has made these process more easy and efficient. I am going to use some of the python packages like plotly to make interactive dashboard and make models that can make great predictions.

Secondly, working with time series forecasting is a crucial part in my dissertation. I have number of different targets in my dissertation. I have been looking for ways to predict the number of suicides in upcoming years. My interest in time series and ML made me dive deep into sophisticated time series models like SARIMA and VAR to make models on the suicide data and forecast future suicides in different countries.

Thirdly, we need a database server for data to be stored on the server. I will be using PSQL or MySQL servers for data storage and management. I want the data in my DB to be updated time to time and my model has to be updated based on the new data injected in each time. The reason for choosing these DB’s is the flexibility of usage and its syntax matching with Structured Query Language (SQL) minute differences.

Fourthly,

**Related work**

ARIMA Model and FBProphet models used for predicting suicide deaths around the world.(Kumar and Susan 2020). Covid-19 was a very sensitive topic in recent year. Many people have affected by it and lost their life. The dataset is very similar to what I have chosen for my suicide prediction as well. In my view This is an excellent model to take inspiration from. The paper talks about countries including India and to understand the patten in deaths happened around the world. Time series is definitely helped to make predictions about coming years. ARIMA and FBProphet models are used and for analysis data has been split into training and testing.



Another study was done on predicting birth (Włodarczyk et al. 2021), study was trying to figure out the preterm births. This study used machine learning algorithms like support vector machine(svm), random forest, K-Nearest Neighbor, and Convolutional Neural Network (CNNs).

Another study I can point out was done on predicting mortality of predicting attributable to cancer in Qingdao, China: (Qi et al. 2021). They have also used ARIMA Model for prediction of deaths. ARIMA model is combination of autoregressive model and moving average model.Another study was done on prediction o exchange rate (Airiti . 2012). Artificial Nueral Network and ARIMA are used to predict the model.

SVM has been used Time series analysis, like (Huang et al. 2017) has clearly studied classification problem on Breast cancer dataset. This study also checked for different kernel function that used in the SVM Classifier. The outcome of their shows that for large scale datasets RBF kernel based SVM ensembles based onboosting perform better than the other classifiers. SVM is first introduced by (Cortes and Vapnik 1995) shown that it’s better for two-group classification problems.

PCA-KNN model is used in (2018) for financial time series prediction, we could use output from sliding window as input for the KNN Model. Principal Component Analysis (PCA) is used in the transformation of the data as well. Suicide dataset will have to undergo above methods to achieve efficiency and accuracy in modeling or achieve optimum results. Empirically, my assumptions on the previous studies may vary along my research but still this literature review on previous studies has help me improve my preparations for the suicide research in achieving my project goals.

Working with Multivariate time series data, I was looking for models which can making predictions on more than one variable. For example, the (Vector Autoregressive Models for Multivariate Time Series 2006) showed me how relevant is VAR (Vector Autoregressive Models) model for the suicide analysis. It also gave me.

My test includes checking seasonality, tends, stationarity and testing statistical models for finding the best model for prediction. AIC and BIC

**References**

*(3) 1/4: What is Streamlit - YouTube*. Available from: https://www.youtube.com/watch?v=R2nr1uZ8ffc [accessed 4 June 2022].

*Airiti Library\_Comparative+Study+of+Artificial+Neural+Network+and+ARIMA+Models+in+Predicting+Exchange+Rate*. Available from: https://www.airitilibrary.com/Publication/alDetailedMesh?docid=20407467-201211-201512080011-201512080011-4397-4403 [accessed 8 April 2022].

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning* [online], 20(3), pp.273–297.

Huang, M.W., Chen, C.W., Lin, W.C., Ke, S.W. and Tsai, C.F. (2017). SVM and SVM ensembles in breast cancer prediction. *PLoS ONE*, 12(1).

Kumar, N. and Susan, S. (2020). COVID-19 Pandemic Prediction using Time Series Forecasting Models. *2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020* [online], 1 July 2020.

Qi, F., Xu, Z., Zhang, H., Wang, R., Wang, Y., Jia, X., Lin, P., Geng, M., Huang, Y., Li, S. and Yang, J. (2021). Predicting the mortality of smoking attributable to cancer in Qingdao, China: A time-series analysis. *PLOS ONE* [online], 16(1), p.e0245769. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0245769 [accessed 8 April 2022].

Tang, L., Pan, H. and Yao, Y. (2018). K-nearest neighbor regression with principal component analysis for financial time series prediction. *ACM International Conference Proceeding Series* [online], 12 March 2018, pp.127–131.

*Vector Autoregressive Models for Multivariate Time Series*. (2006). *Modeling Financial Time Series with S-PLUS®* [online], 9 October 2006, pp.385–429. Available from: https://link.springer.com/chapter/10.1007/978-0-387-32348-0\_11 [accessed 4 June 2022].

Włodarczyk, T., Płotka, S., Szczepański, T., Rokita, P., Sochacki-Wójcicka, N., Wójcicki, J., Lipa, M. and Trzciński, T. (2021). Machine learning methods for preterm birth prediction: A review. *Electronics (Switzerland)*, 10(5).