California State University San Bernardino
School of Computer Science and Engineering

CSE 5953 Undergraduate Independent Study

<u>Date</u>

December 8, 2021

<u>Time</u>

11:30

<u>Location</u>

Jack Brown Hall, room 360

<u>Title</u>

"On the useful of VAERS data for vaccine safety research and development: A case study for machine learning based analysis of post Covid-19 vaccine effects"

<u>Student</u>

James Flora

<u>Advisor</u>

Bilal Khan

<u>Abstract</u>

The Center for Disease Control's (CDC) Vaccine Adverse Event Reporting System (VAERS) was developed and launched in 1990 for monitoring a given vaccination's adverse effects on the general population. Several studies are published on the nature of VAERS as a passive reporting system and utilized its data for vaccine related research and development. Given the recent emergence of the SARS-CoV-2 Coronavirus and an influx of 643,522 distinct VAERS reports (by 11/16/2021), it is critical to highlight the usefulness of VAERS data as well as document the protocols that are required to particularly carry out statistical analyses on VAERS data. Based on the comprehensive data cleaning/preprocessing efforts and the application of advanced machine learning approaches on Covid-19 vaccines, the present study proposes a set of recommendations that are critical to apply machine learning modeling approaches or exploratory analyses of vaccines via VAERS data. While VAERS data is unequivocally useful for vaccine safety research and development, duplicates in VAERS samples (i.e., a report by the same individual divided into multiple rows), spelling differences due to typing mistakes and missing values should be handled carefully as statistical analysis techniques can be greatly impacted by these critical factors. The present case study on VAERS data was accompanied by a customized online survey providing 211 distinct reports of the Covid-19 post vaccination effects. By applying exploratory data analysis and advanced machine-learning approaches (i.e., association rule mining, self-organizing maps, correlation analysis, bipartite graphs, and Bayesian Networks), a comprehensive correlation analysis of the post Covid-19 vaccine effects from both VAERS and online survey data was carried out. Various useful data preprocessing/cleaning techniques were pinpointed which should be considered as part of VAERS and a baseline decision support for Covid-19 vaccine safety was proposed. Additionally, rapidly evolving nature of the Coronavirus vaccine was focused with particular attention devoted to the effects reported in VAERS reports for children (under the age of 16), in order to promote vaccine safety for at-risk groups.