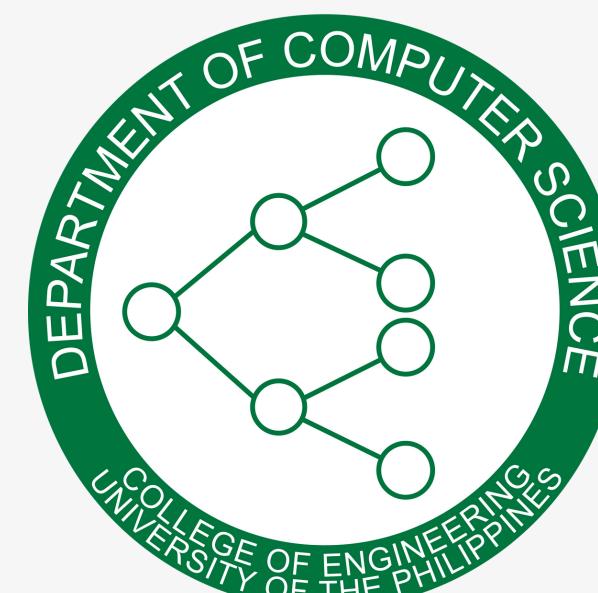


The PhilHealth Queen: Time Series Analysis of the Mis/Disinformation Tweets on PhilHealth Corruption Attributed to Risa Hontiveros

Felix IV Bueno | James Adrian Perez | Clarisse Bianca Bucu || Department of Computer Science
fqbueno@up.edu.ph jvperez@up.edu.ph ccbucu@up.edu.ph info@dcs.upd.edu.ph



PROJECT OVERVIEW

With the world currently being in a period of information and social platforms being easily accessible, mis/disinformation is spreading at a rapid rate. In this study, we look at the trends of frequency of false tweets against Risa Hontiveros as the 2022 election approaches.

NULL HYPOTHESIS: THE AMOUNT OF MIS/DISINFORMATION TWEETS ATTRIBUTED TO RISA HONTIVEROS IS NOT SIGNIFICANTLY DIFFERENT BETWEEN THE TWO PERIODS (ELECTION & PRE-ELECTION)

HYPOTHESIS: THE AMOUNT OF MIS/DISINFORMATION TWEETS ATTRIBUTED TO RISA HONTIVEROS SIGNIFICANTLY INCREASED AS THE 2022 ELECTIONS CAME CLOSER



METHODOLOGY

DATA COLLECTION

A total of 158 tweets were collected with the advanced search feature using keywords like `risa philhealth`

DATA PREPROCESSING

Raw data was preprocessed such that missing values were dealt with, formats were consistent, certain values were interpolated, etc.

DATA EXPLORATION

Several aspects of the data (ex. account types) were visualized prior to the actual statistical testing.

RESULTS & DISCUSSION

Figure 1 below shows the various peaks and change points observed in the data. The peaks range from 5 counts to 15 counts. The highest peaks are observed in the years 2020 to 2022, with 2020 having the highest peak, followed by 2022 (after the exact election), and 2021 having the lowest peak. Change points are also evident in the graph, with a mean shift observed in the year 2017. The standard deviation is greater during the periods of 2019 to 2022 (before the election) and 2022 after the election.

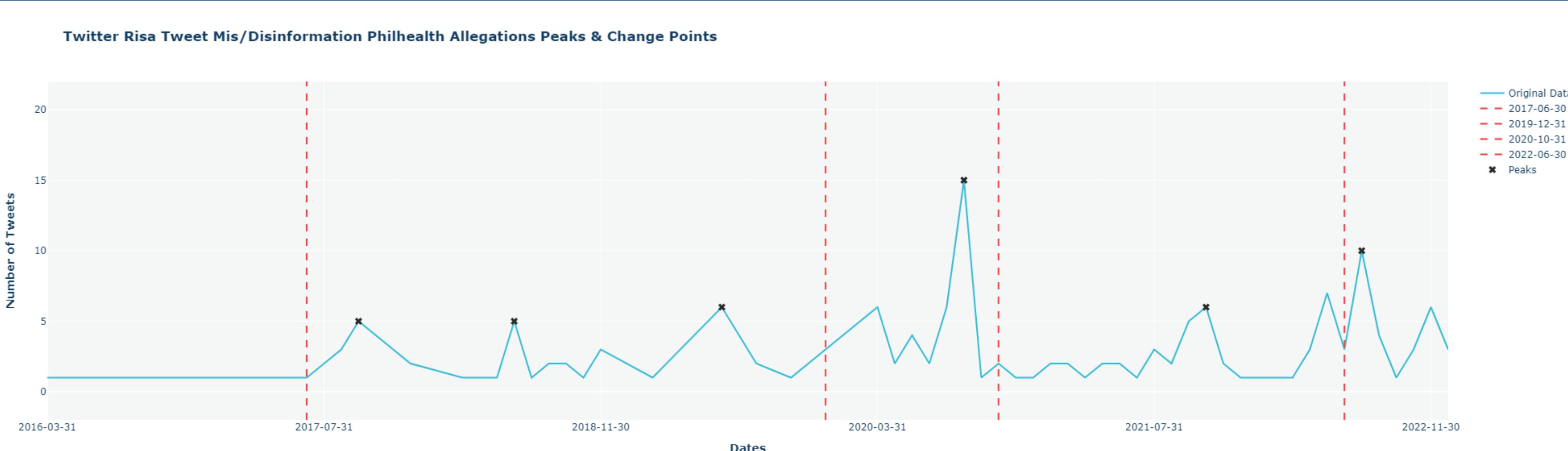


Figure 1. Peaks and Change Points of Mis/Disinformation tweets in relation to Risa's Philhealth Allegations

Figure 2 shows the segmented Linear Regression for the two periods being compared. The p-value for the pre-elec period is **0.0084** which shows that there is a relationship between the two variables. Contrary to this, the p-value for the elec period is **0.3292** which shows that there is no distinct relationship between the two variables during this period. Note also that the standard deviation during this period is higher than the pre-elec period.

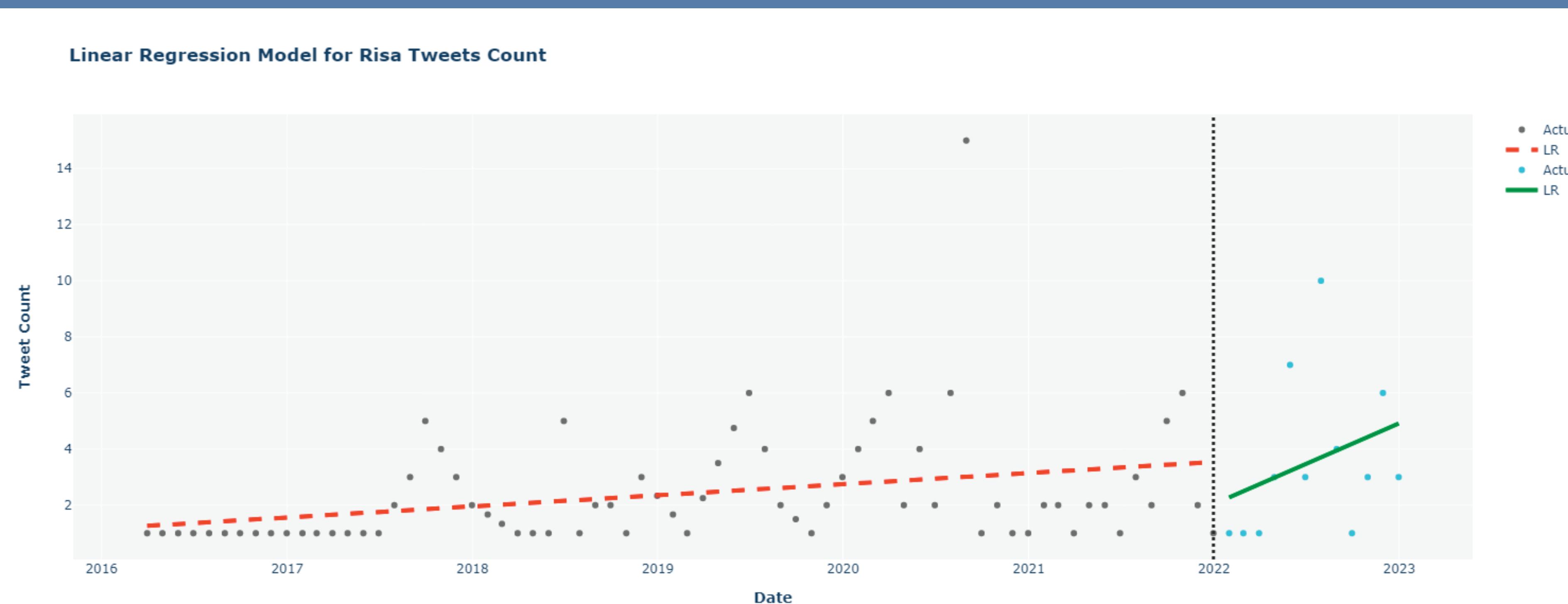
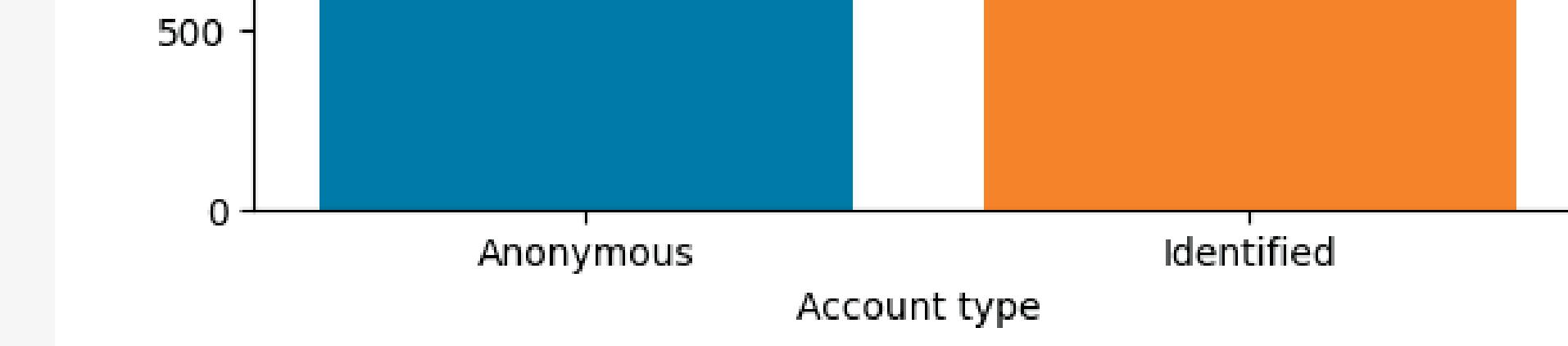


Figure 2. Segmented LR vs SVR for the two periods (pre-elec & elec)

p-value : 0.1321

test statistic : 309.0

Given a significance level of 0.05 and a p-value of 0.1321 after performing the **Mann-Whitney U Test**, the null hypothesis is not rejected. This means that there is no significant difference in the number of mis/disinformation tweets on PhilHealth corruption attributed to Risa Hontiveros between the two time periods (pre-election and election).



STATISTICAL TESTING

A non-parametric test (Mann-Whitney U) was used to test for the significant difference of the two periods (pre-election & election)

DATA MODELING

Tweet frequency was analyzed by performing peak-finding and change point detection

CONCLUSION & FUTURE WORK

In conclusion, The results show no significant difference in the frequency of tweets between the pre-election and election period. To expound on these findings, it is recommended to further explore mis/disinformation tweets regarding this topic to gain a larger dataset.

ACKNOWLEDGEMENTS

Our sincerest thanks to our CS 132 professor **Sir Paul Regonia**, for his exceptional guidance & support throughout our project, and to all our loved ones, for always supporting & believing in us.