Tembo's 2021 Elections Prediction Model

Bernard Tembo, PhD

11th August 2021 – Version 1.0

Introduction

This write-up documents the Tembo's Prediction Model for Zambia's 2021 Elections. It focuses on the Presidential elections. The model is built bottom-up from constituency level. At this time, Zambia (in 2021) has 156 constituencies. All the historical and current registration data used in developing this model is obtained from the Electoral Commission of Zambia (ECZ). **R** programming software (an open source software) was used to simulate different elections scenarios.

The main purpose of the Tembo Model is to predict election outcomes. However, whereas in the past (in 2016), I attempted to predict before hand who the likely winner would be, this model instead focuses on predicting the election winner as results are being announced by ECZ. In addition, this model can be used during election campaigns planning: as key constituencies for all parties involved in an election can be identified easily. To give an example, Malole and Monze Central are strategic constituencies for PF and UPND respectively. Therefore, by using of this Model, parties could better allocate their resources to achieve a desired election outcome. I must say nonetheless that in this year's election, I did not use this model for planning purposes nor did I offer any services using the model. I developed this model out of curiosity and interest to contribute to data analytics in Zambia.

I am aware that people desparately want to know which party (PF or UPND) is likely to win this year's election. However, no one can with certainty give you an answer to this issue, and I too will not attempt to answer this question. I can, with confidence, nonetheless say that the chances of having a **Re-run** are considerably less than the chances of either PF or UPND winning in the first round. Furthermore, based on the analysis of the model outputs, a PF win has distinct characteristics from that of a UPND win. This implies that the two wins cannot be confused with each other.

1 Brief Model Description

Below is the model function that was used to generate the simulated election outcomes data:

$$TV_{sc} = f(VR_c, VTO_{c,sc}, RV_{c,sc}, PV_{c,p,sc}) \tag{1}$$

$$PV_{c,p,sc} = f(AV_{c,sc}, PS_{c,p,sc,h}, \varepsilon) \tag{2}$$

Where:
TV_{sc} is the total votes cast in that scenario (sc);
VR_c is the total number of registered voters in a constituency (c);
$VTO_{c,sc}$ is the voter turnout in a constituency (c) for a particular scenario (sc);
$RV_{c,sc}$ is the total rejected votes in a constituency (c) for a particular scenario (sc);
$PV_{c,p,sc}$ is the total votes a party (p) gets in a constituency (c) for a particular scenario (sc);
$AV_{c,sc}$ is the total number of available valid votes in a constituency (c) for a particular scenario (sc);
$PS_{c,p,sc,h}$ is the historical (h) votes shares a party (p) – more accurately PF and UPND – got in the past elections (2011, 2015 & 2016) in a constituency (c); and
ε is the error factor in party shares in a constituency (c) for a particular scenario (sc).

The data that was generated using the model function described above was cleaned and processed, and fed into a number supervised learning algorithms. The Random Forest algorithm was chosen, this algorithm is at the core of the Tembo Model. You can read more on Machine Learning (ML) here.

2 Model Web Applications

Two mobile phone friendly web applications (Web Apps) have been developed, but currently only one has been published. However, the second and main Web App will be published this Thursday (12th August) before 5pm CAT.

2.1 Provincial level Web App

This Web App is a toy application which takes user input. It aggregates constituency data to provincial level for prediction purposes. A total of 13, 806 election simulations (total of 138,060 observations – 10 provinces multiplied by 13,806 simulations) are captured in the model. As mentioned in the Introduction, I avoided imposing my views on election outcomes. Therefore, these 13,806 simulations are split as follows: PF and UPND (each) win 34% of the time, with the reminder of 32% scenarios resulting in a re-run.

The Provincial level model is published here. Feel free to play with it! Press the **submit** button and also move around the knobs. For the techies, here is a bit of the technical aspects of this provincial model: Model Accuracy is 94% while Kappa is 92%.

2.2 Constituency level Web App

This Web App is the main output of this project. It will take minimal user input, as the aim of this App is to plot and disseminate the predictions based on the official election results as announced by ECZ. Note, only official ECZ results will be fed into the Model – for prediction purposes. These results will be obtained from this ECZ webpage and other reliable sources that will be capturing the ECZ announcements. In addition, I would greatly appreciate if you (the general public) also send through the data that you manage to capture from ECZ – see my contact details below. However, all information from the general public will be crossed checked to ensure correctness of the data.

So how will this work? I have designed the model to have a default setting of 50%-50% chance for a PF and UPND first round win, then of course there is a chance of a re-run (at 25%) – meaning there is a 75% chance that the election will be decided in the first round. A 50%-50% winning chance for PF and UPND is a non-committal design. However, as results start coming in, the model will automatically update the winning probability of each of the two main parties and also show the probability of a re-run. Such a model design implies that even without a complete announcement of constituency results, one can be able to identify which party has merged victorious – in the case of a first round 50%+1 victory. This is the kind of dynamic forecasting you would see on BBC, CNN or sport betting websites. I am excited about this output.

The Constituency level model will be published here. Feel free to embed it into your website – it is published under a Creative Common License (CC BY-NC). A total of 18,936 election simulations (total of 2,954,016 observations – 156 constituencies multiplied by 18,936 simulations) are captured in the model. The simulations are split as follows: PF and UPND (each) win 37.5% of the time, with the remaining 25% scenarios resulting in a re-run. Techies, here is a bit of the technical aspects of the constituency model: Model Accuracy is 94.4% while Kappa is 91.6%.

3 Closing Remarks

This model documentation is on-going.

My email: ben2great@yahoo.co.uk