CS350 Proposal - *Group Y*
Members: Nicholas Bosworth 20186504, Ludovic PRIGENT 20186520, Steven Hans Limantoro 20186913; Luc MICHELS 20186478, Eric WILDERS 20186404
App Name: KAIST Analytica
App Logo:

Target Domain: Social, Tools, Simulation
Expected Users: Australian political parties, policy institutes and newspapers
Description:
This app aims to predict the results of Australian state elections through supervised machine learning techniques applied to localised Google Trends data. The app's creation was motivated by the upcoming state elections in Australia, which will be held in November 2018 and March 2019.

Traditional polling services have struggled to accurately report information in a more polarised political environment, as seen in the inaccurate predictions during the 2016 US election and 2016 Brexit vote in the United Kingdom.

By providing a more modern method of predicting election results this app intends to provide an alternate and more accurate polling model for Australian political parties and policy institutes.

Data Summary:
This project relies upon data sourced from Google Trends and Australian State Electoral Commissions. The project intends to compare Google Trend data and State Election results for all elections since 2004.

The scope of this project is limited by the availability of Google Trends data, which is only recorded from 2004 in Australia.

The project is also limited by the locations being investigated. The project will be sourcing information for all 8 of Australia's states and territories. These regions have different electoral rules and term periods, so the dates used to retrieve google trend data will need to change according to the location being investigated.

Google trends provides information on the popularity of search terms in a given time period and area by showing the frequency in search volume as a proportion of Google's total search volume.

The project will use Google Trend data to investigate the most popular search terms in each Australian state in the months leading up to an election.

The project will make sure to test the integrity and consistency of the Google Trend data because there are known series breaks due to algorithm changes in 2011 and 2016. Google changed its geographical assignment algorithm on the 1/1/2011, and changed their data collection method on 1/1/2016.

Each location has its own electoral commission which digitally stores the results of each election. The election data is reported according to polling booth, electoral district and statewide result. The project is primarily concerned with the statewide result because it is consistent with the location used to search Google Trends data.

Each state electoral commission stores its data in different file formats, these include both machine-readable and not machine readable formats. This project will need to manually record some data to make it useable.

The majority of states in Australia have maintained consistent data collection throughout the time period being investigated. However, Queensland changed the structure of its parliament in 2015, so the integrity of its data will need to be tested.

Data Sources:

Google Trends:
https://trends.google.com/trends/?geo=AU
A dataset provided by Google summarising requests using the google search engine which is updated in real time.

Australian electoral websites - summarized here:
http://australia.census.okfn.org/dataset/election
This website was created by Open Knowledge International on the 21st Feb 2015 and provided with official data from the Australian government. This project will be using the election result data provided by each state's electoral commission.

Data Usage/Access Methods:

https://www.npmjs.com/package/google-trends-api
This API will permit us to retrieve current google search trends related to specific political terms using a nodejs server.

For the Australian election results we will manually download the raw census datasets from the following link: http://australia.census.okfn.org/api/entries.json

This project will use TensorFlow as the framework for a supervised machine learning program that uses the datasets to predict future election results.

Suitability of the data: 4 (somewhat certain)
The team has some experience with web applications, using nodejs and json parsing. However, there is a lack of experience with API usage and machine learning techniques.

Functional Requirements		
Requirement Name	**Description**	**Priority**
Voting outcome	The predicted vote outcome for the two main parties in a specific state will be displayed	High
State selection	A specific state can be selected from a drop-down menu	High
State colour	A colour relating the results will be displayed	Medium
Voting Percentage	The predicted percentage vote for the two-party-preferred result (percentage of votes received by the two major parties after vote preferences are distributed)	Low
Comparison with actual results	If a prediction for an election that has already occurred is asked, the real results will be displayed as well for comparison	Low

Non-Functional Requirements		
Requirement Name	**Description**	**Priority**
Availability	The service will be up 100% of the time	High
Effectiveness	The system's prediction will be within 10% of actual results	High
Response time	Within 3 seconds of the request	High
Robustness	The machine learning algorithm will have a stable performance	High
Compatibility	The application will be available on at least the following browsers: Google Chrome, Mozilla Firefox, Safari	Medium
Usability	A graphical user interface facilitating the usage for our target audience.	Medium
Open-source	The code will be accessible online on github	Medium
Scalability	Our system will be able to take requests for multiple countries	Low