A Predictive Government Decision Based on Citizen Opinions – Tools & Results

Mohamed Adel Rezk
Insight Centre for Data Analytics
National University of Ireland, Galway
mohamed.adel@insight-centre.org

Adegboygga Oj0
Insight Centre for Data Analytics
National University of Ireland, Galway
adegboygga.ojo@insight-centre.org

Ghada A. El Khayat
Department of Information Systems and Computers
Alexandria University
ghadaek@gmail.com

Safaa Hussein
Department of Information Systems and Computers
Alexandria University
safaa26bedo@yahoo.com

ABSTRACT

Research on citizen satisfaction with respect to public policies has significant public and political value. Politicians are generally seeking effective public policies that favourably impacts citizens’ satisfaction. Citizen satisfaction index is a plausible mechanism for public policy makers to monitor and evaluate the public policies. While surveys on citizen satisfaction are common among public policy makers to monitor and evaluate the public policies. Given that surveys and evaluations related to citizen satisfaction are retrospective, remedial actions when necessary are always somewhat late. We describe in this poster a predictive analytics framework for citizen satisfaction with respect to public policy based on the previous citizen sentiments past related policies.

CCS CONCEPTS
- Applied Computing → E-government;

KEYWORDS
Government Decision Support; Decision Analytics; Policy Acceptance Prediction; Citizen Satisfaction; Policy Aspects; Opinion Mining; Sentiment Analysis; Semantic Relatedness; Topic Modeling; Social Media; Unstructured Text Analysis

ACM Reference format:
DOI: 10.1145/3209415.3209504

1 INTRODUCTION

In our earlier work, we described a Government Decision Analytics Framework (Gov-DAF) [2, 6] which includes a Citizen Satisfaction Analysis (CSA) pipeline shown in Figure 1 and elaborated in [1]. Gov-DAF addresses the problem of lack of tools to support critical government decisions regarding the public policy-making process. In the framework, the knowledge base comprises aggregated citizen opinions expressed on social media platforms such as Twitter. We report here some details of the citizen satisfaction analysis (CSA) pipeline implementation as well as the results obtained. The CSA pipeline was implemented by integrating state-of-the-art unstructured data analysis tools and techniques. Specifically, the pipeline algorithm/logic was implemented using python2 and topic modeling was carried out using gensim’s LDA3. Data persistence was implemented using NoSQL storage and retrieval using mongodb4. Results from the implementation is reported in Experimentation section of this paper and illustrated in Figure 3 and Figure 4. In the concluding section, we highlight the CSA pipeline limitations and future research opportunities.

1 Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

2 https://www.python.org/
3 https://radimrehurek.com/gensim/
4 https://www.mongodb.com/
2 EXPERIMENTATION

2.1 Experiment Details

We conducted an experiment to test the pipeline described in Figure 1. The experiment started with feeding the CSA pipeline a new public policy “20-Year Strategy for the Irish Language 2010-2030”\(^5\) required for examination in pdf format. Then the pipeline ingestion component extracts and cleanses the input policy text to build its textual representation. Then that representation is passed to the Latent Dirichlet Allocation (LDA)\(^6\) topic-modeling algorithm to identify the topics within the public policy text. Following this, a topic co-occurrence matrix is build to represent the policy document. These topics are then used to filter tweets related to the policy text. In knowledge base, each of the topics is associated with the tweets or micro-blogs. Finally, we apply a machine learning based sentiment analysis to compute the sentiments over the stored tweets. The resulting knowledge base is subsequently used for generating or predicting the citizen satisfaction index for the input policy document.

2.2 Sample Results

We describe below sample screens of the resulting knowledge base. In Figure 3 we see sample satisfaction matrix records containing co-occurred terms \([A, B]\), their related tweet ids, related tweets classified to negative, positive, and neutral, and finally the summation of those classifications.

In Figure 4 we focus on the tweets representation and classification part, here we store only tweets ids after classifying the terms \([A, B]\) related tweets to positive, negative and neutral.

3 CONCLUSION

As shown in experimentation section, the CSA pipeline is used to measure the satisfaction of citizens towards public policies. This can be done either for a current policy or a policy under development. CSA pipeline results accuracy is tightly coupled with two factors, First, the micro-posts dataset size and quality, where quality means its relatedness to the public policy under analysis. Second, the prediction algorithm applied on the built matrix to estimate citizen satisfaction for public policies still under development.


\(^6\) https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation
REFERENCES


