WEEK 8: FINAL VERSION OF PROJECT PROPOSAL

Lasya Priya Konduru

School for Professional Studies,

Saint Louis University

07-16-2023

Instructor – *Dr. Srikanth Mudigonda*

1. **INTRODUCTION**

* *Describe the problem you intend to address.*

Problem: **Investigating the Factors Affecting News Popularity in Multiple Social Media Platforms.**

This research aims to analyze the dataset to determine the primary characteristics that contribute to the popularity of news stories and the amount of engagement they receive across various social media platforms. By examining the features and variables provided in the dataset, you can explore the connections between various factors and the popularity metrics of news articles.

* *Describe specific questions (2-3) that a solution that you would be proposing will help in answering.*

The proposed solution may help in addressing the following questions:

1. **Which aspects of articles are most responsible for their virality on various social media sites?**

The value of many aspects, such as article content, title, publisher, publishing time, or user interaction metrics, in determining the popularity of news can be determined through an analysis of the dataset.

1. **To what extent does the level of interest in each article on social media reflect the sentiment analysis of user comments and interactions with that story?**

By using sentiment analysis of the provided data, we can gain insight into how users' positive or negative emotions influence the popularity of articles.

1. **Is it true that certain types of articles always do better than others across all major social media sites?**

We can see if some themes are more likely to generate popularity and engagement across different platforms by evaluating the dataset and classifying news stories based on their topics or domains.

By addressing these issues, the suggested approach can reveal trends in widely discussed news stories, illuminate the connection between news audience sentiment and audience involvement, and reveal insights into what drives news popularity on social media. These results help content producers, publishers, and social media sites fine-tune their methods, increase their audiences' participation, and broaden the reach and effect of their articles.

* *Who will be the decision-makers and/or users of your solution. Describe their role.*

Depending on the product and context, the decision-makers and users may vary. The following is a list of potential stakeholders and the responsibilities they play:

1. **News Publishers and Content Creators**: News publishers and content creators play a big part in making and spreading news articles. They can use the information from the research to improve their content strategies, figure out what makes news popular, and make intelligent choices about article titles, when to publish, how to format content, and what to write about. Analysis can help them change their content to get more people interested and reach a bigger audience on social media platforms.
2. **Social Media Platform Managers**: Algorithm and content curation managers can utilize analytics to improve their algorithms, news suggestions, and user experience. The information they learn can help them figure out what makes news stories popular and teach the platform's algorithms how to show users relevant and interesting content.
3. **Marketing and Advertising Professionals**: The analysis can help marketers and advertisers determine which news stories and subjects are hot on social media. They can use these insights to align their marketing strategies better, choose the right news stories for their marketing campaigns, and better reach out to specific groups of people.
4. **Researchers and Academics**: Researchers and academics in social media analysis, media studies, and data science can use the analysis results to contribute to academic research, create new models or algorithms, and learn more about how news becomes popular on social media platforms.
5. **General Public and News Consumers**: The analysis can improve news content, suggestions, and social media relevance for the public and news consumers. By making it easier to understand what makes news famous, the solution can help people find more interesting, reliable, and varied news stories.
6. **DESCRIPTION OF PROPOSED SOLUTION**

* *Describe in detail your proposed solution. Provide appropriate citations that support your choice of a solution.*

The proposed solution combines sentiment analysis, topic modeling, and machine learning to discover what makes news popular and get people to talk about it on social media platforms.

1. Sentiment Analysis: Sentiment analysis will classify how people feel about news articles in user-generated material. This study will help you understand how people's feelings spread and how that affects how popular news is. Pang and Lee (2008) say that sentiment analysis can be done with lexicon-based methods, machine learning classifiers, or deep learning models.
2. Topic Modeling: Topic modeling algorithms like Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003) will be used to find the underlying topics and themes in the news stories. This will make it easier to put news stories into groups and determine how popular specific topics are on different social media sites.
3. Machine learning: Machine learning will be used to predict news stories' popularity based on the dataset's traits. For example, regression or classification models will be used. This can help determine the most important things that interest people in the news and give you information you can use to make decisions.

The suggested solution will use existing research and methods in sentiment analysis, topic modeling, and machine learning to look at the dataset and determine what makes news popular on social media platforms. Several research studies by Agarwal et al. (2011), Blei, Ng, and Jordan (2003), and Pang and Lee (2008) have shown that these methods work well for analyzing sentiment, finding subjects, and predicting popularity in social media settings.

* *Describe one or two alternatives to your proposed solution.*

1. Trend Analysis and Recommendation System: Instead of trying to predict how popular news will be, a different option is to make a system for analyzing trends and making suggestions. This system would find new news topics, keep track of how popular they are over time, and give users personalized suggestions based on their hobbies and patterns of engagement. It would help people keep up with what's happening in the news and find relevant and interesting material on social media platforms.
2. Content Optimization and A/B Testing: Another option would be to do A/B tests with different versions of news stories to figure out how to make their content work best on social media platforms. With this method, you would try out different article titles, formats, or visual features in a planned way to find the most exciting and popular combinations. By looking at the results of these trials, decision-makers can learn more about how to optimize content and make changes to how they create content.

These alternatives offer different ways to deal with the problems and goals related to popular news and how people interact with it on social media sites. Each option has its pros and cons, such as personalized suggestions, content optimization, and A/B testing, which decision-makers can weigh based on their goals and available resources.

* *Explain why your proposed solution is as good as, or better than, the alternatives you have considered. Here, too, you should provide appropriate evidence to support your answers.*

The proposed method is just as beneficial as, even better than, the alternatives since it utilizes multiple approaches to understanding what makes news popular, including sentiment analysis, topic modeling, and machine learning.

By analyzing sentiment, identifying topics, and predicting popularity, decision-makers can optimize content strategies and provide personalized recommendations. The suggested approach uses prior studies in sentiment analysis, topic modeling, and machine learning to produce valuable outcomes.

* *Explain why your proposed solution will be viable (i.e., can be implemented in a way that will help you address the problem).*

Because it uses previously developed approaches and research in sentiment analysis, topic modeling, and machine learning that have already been effectively deployed in social media environments, the proposed solution has a good chance of being implemented. The availability of the dataset titled "News Popularity in Multiple Social Media Platforms" offers a solid basis for implementing and evaluating the solution.

* *Explain how your proposed solution will meet the needs of the decision-makers/users of your solution.*

The proposed solution corresponds to many decision-makers requirements, such as marketing and advertising professionals, content providers, social media platform management, researchers, and the general public. It offers insights into the factors that drive news popularity, optimizes content strategies, improves algorithms and suggestions, and assists users in discovering news items that are pertinent to them and enjoyable to read.

* *Describe any potential constraints that may affect successful implementation of your proposed idea.*

Investigating news popularity on multiple social media platforms may be constrained by several variables. These are a few key constraints:

1. Data Availability and Quality: The proposed solution's execution depends on dataset availability and quality. The "News Popularity in Multiple Social Media Platforms" dataset may have size, representativeness, or diversity issues. Insufficient or skewed data may influence conclusions' reliability and generalizability. This constraint necessitates rigorous data collection and preparation to produce a broad and diversified dataset that accurately depicts the target social media networks and their news content.
2. Computational Resources and Scalability: Large-scale sentiment analysis, topic modeling, and machine learning require computational resources. Processing and interpreting a lot of textual data can be resource intensive. Implementing the proposed approach may require high-performance computer clusters or cloud-based resources. Optimization, parallel processing, and distributed computing frameworks can improve solution scalability.
3. Data and machine learning expertise: Implementing the notion needs data analysis, sentiment analysis, subject modeling, and machine learning skills. Applying these methods to the dataset requires experts in natural language processing, statistical analysis, and machine learning algorithms. A lack of expertise or resources may hinder results interpretation and predictive model creation. Domain specialists, data scientists, sentiment analysis, and machine learning researchers might be consulted to optimize resource use.

* *Explain how you intend to address these constraints.*

The following strategies can be employed to address potential constraints in executing the proposed idea:

1. Data Availability and Quality:

* **Data Collection**: Collect a broad dataset that spans several social media platforms and a wide range of news items. Data can be collected from multiple sources, APIs, and web scraping.
* **Data Preprocessing**: Rigorous preprocessing steps should be implemented to ensure data quality. This may involve removing duplicates, handling missing values, addressing noise, and cleaning textual data using text normalization, stop word removal, and special characters or emojis.
* **Data validation**: It ensures dataset reliability and representativeness. This may involve cross-referencing data with other sources or sample validations to ensure dataset accuracy and consistency.

1. Computing Resources and Scalability:

* **Resource Optimization**: Feature engineering, dimensionality reduction, and parallel processing can optimize computing resources. These methods simplify sentiment analysis, topic modeling, and machine learning algorithms.
* **Cloud-based Solutions**: Leveraging cloud-based platforms, such as Amazon Web Services (AWS) or Google Cloud Platform (GCP), can provide access to scalable computing resources, allowing researchers to dynamically adjust their computational needs based on the size and complexity of the dataset.
* **Distributed Computing**: Apache Spark can divide computational effort across a cluster of servers, making large-scale dataset processing faster and more efficient.

1. Data Analysis and Machine Learning Expertise:

* **Collaboration and Skill Enhancement**: Working with domain experts, data scientists, or researchers in sentiment analysis, topic modeling, and machine learning can help. Collaboration can involve knowledge-sharing, brainstorming, and using experience to solve problems.
* **Skill Development**: Researchers can enhance their skills by taking advantage of online courses, tutorials, and resources available in data analysis, natural language processing, and machine learning. This can entail studying Python or R, comprehending algorithms, and learning the latest sentiment analysis and topic modeling methods.

Through thorough data collection, preparation, optimization of computational resources, and collaboration with domain experts, researchers can avoid potential problems and ensure a more successful execution of the suggested idea. The techniques will be more effective at overcoming these constraints if monitored, validated, and adjusted during the research process.

**REFERENCES**

* *Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. In Proceedings of the workshop on languages in social media (pp. 30-38).*
* *Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3, 993-1022.*
* *Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.*