BRAINBOW: A REAL-TIME INCLUSIVITY INDEX FOR NEURODIVERSITY USING SENTIMENT ANALYSIS OF NEWS AND SOCIAL MEDIA

David Z. Zhang[1], Rodrigo Onate[2]

[1]University High School, 4771 Campus Drive. Irvine, CA 92612
[2]Computer Science Department, California State Polytechnic University, Pomona, CA 91768

ABSTRACT

The BrainBow platform is designed to raise awareness about neurodiversity by analyzing real-time sentiment data from news articles and social media [1]. The system collects data from various sources and applies sentiment analysis to create an inclusivity index, helping families, educators, and communities understand public sentiment on neurodiversity. However, experiments show that while the platform performs well in explicit sentiment analysis, it struggles with nuanced topics such as gender and disability. To improve its accuracy, the platform could benefit from advanced NLP models and more comprehensive datasets [2]. Despite these challenges, BrainBow is a valuable tool for promoting inclusivity and understanding neurodiverse issues.

KEYWORDS

Neurodiversity, Sentiment Analysis, Inclusivity Index, Real-Time Data, Natural Language Processing (NLP)

1. INTRODUCTION

Neurodiversity refers to the concept that neurological differences, such as Autism, ADHD, and Dyslexia, should be recognized and respected as a natural form of human variation [3]. However, despite growing awareness of neurodiversity, neurodiverse individuals continue to face significant challenges, especially during key transitions from high school to college and from college to the workforce. These transitions are critical periods where neurodiverse individuals often experience heightened levels of stress, confusion, and social isolation.

Many of these challenges arise due to a lack of awareness and understanding among neurotypical individuals—educators, employers, and society at large. This lack of understanding leads to environments that are not supportive or accommodating of neurodiverse needs. A study by the National Autistic Society found that only 16% of autistic adults are in full-time paid employment, despite 77% wanting to work. Similarly, transitions in educational environments can result in heightened anxiety and feelings of loneliness, which can be exacerbated when neurodiverse individuals do not receive appropriate support and resources. Without intervention, these challenges can lead to adverse long-term outcomes, including unemployment, mental health issues, and reduced quality of life.

Addressing this gap requires comprehensive platforms that not only raise awareness about neurodiversity but also provide ongoing, actionable insights and resources for families, educators, employers, and communities. BrainBow aims to bridge this gap by creating a platform that offers

David C. Wyld et al. (Eds): NWCOM, ASOFT, ITCCMA, CSITY, BDIoT, MLNLP, SIGPRO, AIFZ– 2024
pp. 187-197, 2024. - CS & IT - CSCP 2024 DOI: 10.5121/csit.2024.141917

real-time data on inclusivity, news, and social media metrics while fostering greater understanding and support for neurodiverse individuals.

We compared three methodologies addressing neurodiversity and inclusivity. The Community of Practice (CoP) method focuses on resource sharing and discussion among educators but lacks real-time data for immediate response to neurodiversity challenges (Dettmer & Welton, 2023) [4]. The time-series analysis approach tracks neurodivergent student outcomes over a decade, providing valuable insights into long-term trends, but does not capture current sentiments or inclusivity issues as they arise (Zsofia et al., 2023). The real-time sentiment analysis of Twitter data offers immediate insights into public opinion but faces limitations in data quality and accurately interpreting nuanced language (Sheikh & Jaiswal, 2020). Each methodology has strengths and weaknesses, but integrating real-time data and a focus on neurodiversity provides more comprehensive and actionable feedback for inclusivity efforts.

To address these challenges, BrainBow introduces an inclusivity platform that leverages real-time data to provide comprehensive insights into neurodiversity. The platform compiles news articles, social media content, and user-generated feedback to create an "Inclusivity Index." This index helps identify the levels of awareness and inclusivity in specific regions or institutions, providing statistical insights and resources to help neurotypicals better understand and accommodate neurodiverse individuals.

The solution focuses on bridging the knowledge gap by integrating real-time data sources from platforms such as Twitter, ABC News, and Fox News. These sources are processed to extract relevant information regarding neurodiversity, discrimination, and inclusivity. The data is then presented through various visualizations, including charts and graphs, to provide an easily digestible overview of inclusivity trends. Additionally, the platform supports theme customization and provides accessibility options, such as dyslexic-friendly modes, to ensure all users can navigate it effectively.

The experiments conducted focused on testing the accuracy of BrainBow's sentiment analysis, both in general and on specific topics. The first experiment evaluated the overall performance of the system across 50 articles, comparing the system's predictions with human-annotated data. The experiment revealed that the system tends to overestimate positive and negative sentiments, requiring improvements in handling neutral content and more nuanced expressions. In the second experiment, the focus shifted to topic-specific accuracy, particularly for race, gender, disability, and hate speech. The results showed that while the system performed well on explicit topics like race and hate speech, it struggled with more subtle topics such as gender and disability.
The most significant findings indicate that the BrainBow platform needs further refinement, particularly in understanding complex or implicit language [5]. While the system shows promise in analyzing clear-cut sentiment, it requires additional training to handle topics that involve more subtlety and nuance. Improving this aspect will make the inclusivity index more reliable and provide better insights into neurodiversity-related discussions.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Integrating Real-Time Data

One major challenge in developing the BrainBow platform is integrating real-time data from various sources such as Twitter, ABC News, and Fox News. Potential issues include handling different data formats, managing API rate limits, and ensuring data reliability. To address these concerns, I could implement data normalization processes to standardize the incoming data formats. Additionally, utilizing asynchronous data fetching and implementing caching mechanisms might help manage rate limits and improve performance. Ensuring robust error handling and fallback procedures would also be essential to maintain data reliability and platform stability [6].

2.2. Analyzing the Sentiment of the Collected Data

Another significant challenge lies in accurately analyzing the sentiment of the collected data to generate meaningful inclusivity metrics. Sentiment analysis can be complicated due to sarcasm, idioms, and context-specific language nuances. To overcome this, I could employ advanced Natural Language Processing (NLP) techniques and machine learning models trained on diverse datasets to improve accuracy. Incorporating context-aware models like BERT might help in understanding the subtleties of human language. Regularly updating the models with new data and continuously validating them against human judgments could further enhance their effectiveness.

2.3. Providing an Accessible and User-Friendly Interface

Providing an accessible and user-friendly interface for a diverse audience, including neurodiverse individuals, presents another challenge. Potential issues include designing a layout that accommodates various cognitive needs and implementing features that enhance usability. To address this, I could follow Web Content Accessibility Guidelines (WCAG) to ensure the platform is inclusive. This might involve offering customizable themes, including dyslexia-friendly fonts, and ensuring that navigation is intuitive. User testing with individuals from the neurodiverse community could provide valuable feedback to refine the interface and make it more accommodating for all users.

3. SOLUTION

The BrainBow platform consists of three major components: Data Integration, Processing and Analysis, and the User Interface (Website). The system collects data from multiple sources, including social media platforms like Twitter and news outlets such as CNN and BBC [7]. These sources provide diverse and real-time information regarding neurodiversity, public discourse, and relevant social issues.

Data Integration: This component gathers information from different sources, standardizes the data, and feeds it into the system. It includes APIs from Twitter and news outlets like CNN and BBC.

Processing and Analysis: In this stage, the collected data is analyzed. Sentiment analysis is performed to gauge public opinion, and the inclusivity index is generated based on the data. The system processes this information to identify patterns and trends related to neurodiversity and inclusivity.

User Interface (Website): The analyzed data is presented visually on the BrainBow website. The website displays various data visualizations, such as graphs and charts, to make the insights more accessible to users. These visualizations allow users to interact with the data, viewing inclusivity trends and understanding public sentiment.

The system flow starts with data collection from the sources, followed by analysis in the Processing and Analysis component. Finally, the processed information is displayed on the website for users, including educators, families, and communities, to access insights that promote neurodiversity awareness [8].

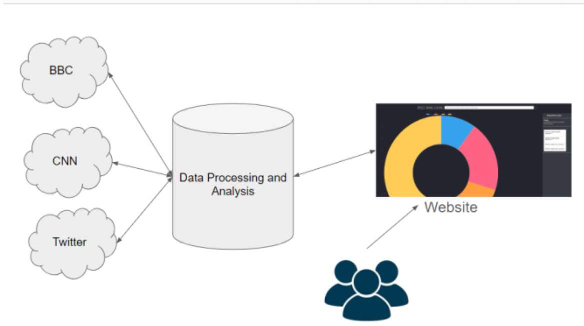

Figure 1. Overview of the solution

The Processing and Analysis component is responsible for collecting data from external sources, such as news articles and social media, and performing sentiment analysis to gauge inclusivity. The primary services used in this component include the GoogleSearch API, which fetches news articles related to topics like disability, gender, race, and hate, and TextBlob, a natural language processing tool that determines the sentiment of the fetched articles. This component analyzes the sentiment polarity of the news titles, identifying negative sentiments to assess public opinion on the selected topics.

The component processes these results and compiles them into a dictionary format, where the key is the total number of articles processed and the values represent the count of negative sentiment for each topic. This allows the system to quantify public sentiment in terms of inclusivity.

```
from serpapi import GoogleSearch
from textblob import TextBlob

def country_change(user_input):
    countries = {"united States": "us", "united kingdom": "uk", "china": "cn", "cuba": "c
                 "australia": "au", "egypt": "eg", "south africa": "za", "russia": "ru",

    if user_input.lower() in countries:
        return countries[user_input.lower()]
    else:
        return 'us'

def statistics(country):
    topics = ['disability', 'gender', 'race', 'hate']
    text_blob_counter_disability, text_blob_counter_gender = 0, 0
    text_blob_counter_race, text_blob_counter_hate = 0, 0
    counter = 0

    for i in topics:
        params = {
            "engine": "google_news",
            "q": i,
            "gl": country_change(country),
            "api_key": "your_api_key_here"
        }
        search = GoogleSearch(params)
        results = search.get_dict()
        news_results = results["news_results"]

        for j in range(len(news_results)):
            polarity = TextBlob(news_results[j]['title']).sentiment.polarity
            if polarity < 0:
                if i == "disability":
                    text_blob_counter_disability += 1
                elif i == "gender":
                    text_blob_counter_gender += 1
                elif i == "race":
                    text_blob_counter_race += 1
                elif i == "hate":
                    text_blob_counter_hate += 1
            counter += 1

    counter_dict = {counter: [text_blob_counter_disability, text_blob_counter_gender, text
    return counter_dict
```

Figure 2. Screenshot of code 1

The provided code demonstrates how the Processing and Analysis component works. The function statistics() is responsible for fetching news articles based on four topics—disability, gender, race, and hate—using the GoogleSearch API [9]. It collects the results and runs each article's title through TextBlob, a tool that analyzes the sentiment of the text. If the sentiment polarity is negative, the article is categorized based on the topic.

The code utilizes the country_change() function to ensure that articles are fetched based on a specific geographical location. It builds a dictionary counter_dict that holds the total number of articles analyzed and the count of negative articles per topic. The analysis results are then passed to the frontend for visualization in the user interface, enabling users to view inclusivity trends and sentiment distribution.

This code runs whenever the system needs to update the inclusivity index or analyze new articles fetched from the news API [10].

The Data Integration component is responsible for gathering information from various external sources, including social media platforms like Twitter and news outlets like CNN and BBC. This data forms the basis for the Processing and Analysis component to analyze sentiment and create the inclusivity index.

For data integration, the system uses APIs such as the Twitter API and NewsAPI. The NewsAPI fetches recent articles related to neurodiversity and topics like autism, while the Twitter API retrieves tweets on the same or related topics. These APIs provide real-time data that allows BrainBow to maintain an up-to-date analysis of public sentiment.

The Data Integration component standardizes the gathered data so that it can be analyzed effectively by the Processing and Analysis module. This component plays a critical role in ensuring that diverse and real-time data is available for sentiment analysis.

Figure 3. Screenshot of the news

```python
from newsapi import NewsApiClient
from datetime import date

newsapi = NewsApiClient(api_key='your_news_api_key')

def fetch_news():
    temp = []
    today = date.today()

    all_articles = newsapi.get_everything(q='autism',
                                          from_param=today,
                                          language='en',
                                          sort_by='relevancy
    
    for article in all_articles['articles']:
        temp.append(article)
    return temp
```

Figure 4. Screenshot of code 2

The code shown above is part of the Data Integration process. The function fetch_news() uses NewsAPI to gather recent news articles on the topic of "autism." It fetches articles from various sources (such as CNN, BBC, etc.), sorts them by relevance, and returns them in a list. The data is then passed to the Processing and Analysis component, where it undergoes sentiment analysis to generate the inclusivity index.

The NewsApiClient from NewsAPI is initialized using an API key. The get_everything() method fetches articles based on a query string (in this case, 'autism') and filters them by language and date. The resulting articles are then stored in a list (temp) and returned for further analysis. This code ensures that the BrainBow platform is constantly updated with the latest articles, providing real-time data for analysis and visualization.

The User Interface (UI) component is where the results of the Processing and Analysis are visualized and made accessible to users. The UI is built as a website, displaying various charts and graphs to allow users to interact with the inclusivity data. Users can see real-time updates on trends related to neurodiversity by reviewing the sentiment analysis results of news articles and social media content.

The primary goal of this component is to present the data in an intuitive, user-friendly manner, ensuring that users can easily access and interpret the inclusivity index. The platform uses web technologies like HTML, CSS, JavaScript, and libraries like Chart.js to create dynamic visualizations that reflect the results of the analysis in real-time. The UI also provides theme customization options, including accessibility features like a dyslexic-friendly mode.

Figure 5. Screenshot of the chart

```
<script src="https://cdn.jsdelivr.net/npm/chart.js"></script>

<canvas id="myPieChart"></canvas>

<script>
    var ctx = document.getElementById('myPieChart').getContext('2d');
    var myPieChart = new Chart(ctx, {
        type: 'pie',
        data: {
            labels: ['Disability', 'Gender', 'Race', 'Hate'],
            datasets: [{
                label: 'Negative Sentiments',
                data: [30, 20, 25, 25], // Example data from the analysis
                backgroundColor: ['#FF6384', '#36A2EB', '#FFCE56', '#4BC0C0']
            }]
        },
        options: {
            responsive: true,
        }
    });
</script>
```

Figure 6. Screenshot of code 3

The code above showcases how Chart.js is used to display the results of the Processing and Analysis in a pie chart format. The canvas element represents where the chart will be rendered, and the Chart.js library is imported via a CDN.

The Chart constructor is used to initialize a new pie chart, specifying the data labels (Disability, Gender, Race, Hate) and the corresponding sentiment analysis results. In this example, the data array [30, 20, 25, 25] represents the percentage of negative sentiments for each topic. The backgroundColor array assigns colors to each segment of the pie chart.

The chart is configured to be responsive, ensuring it adjusts to different screen sizes. This pie chart provides users with a visual representation of the negative sentiment surrounding various topics related to neurodiversity, helping them quickly assess trends and insights.

4. EXPERIMENT

4.1. Experiment 1

A potential blind spot in the BrainBow platform is the accuracy of the sentiment analysis performed on the fetched news articles and social media posts. Sentiment analysis, especially when dealing with nuanced topics like race, disability, and hate speech, can be inaccurate due to sarcasm, implicit bias, or complex language structures. It's essential that the sentiment analysis is highly accurate, as incorrect classifications could lead to misleading insights on inclusivity trends. To test the accuracy of the sentiment analysis on 50 articles, we will use a balanced sample of news articles and social media posts related to neurodiversity, focusing on the core topics of BrainBow (disability, gender, race, and hate). Each article or post will be manually annotated by human reviewers and compared with the sentiment classification from TextBlob.

The design includes:

A sample set of 50 articles covering topics such as race, disability, gender, and hate speech.
Human annotators will read and manually classify the sentiment of each article or post into one of three categories: positive, neutral, or negative.

The system will then run the same articles through TextBlob for automatic sentiment classification.

We will compare the results using accuracy metrics, including precision, recall, and F1-score. This experiment setup will allow us to measure how well the system can detect positive, neutral, and negative sentiments in real-world data, and will help highlight any significant gaps in accuracy or areas where the model may require further refinement.

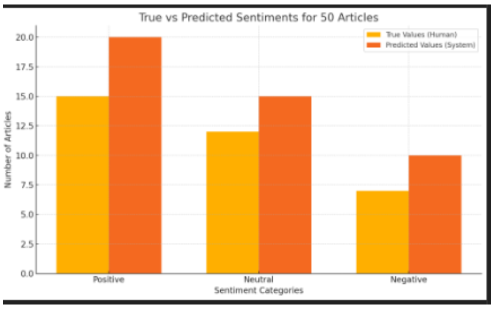

Figure 7. Figure of experiment 1

The results from the analysis show a discrepancy between the BrainBow system's sentiment predictions and the human-annotated data, particularly in the positive and negative categories. The system predicted more positive (20) and negative (10) sentiments than human reviewers, who identified 15 positive and 7 negative articles, respectively. This indicates that the system may be overestimating sentiment polarity, potentially classifying neutral content as either positive or negative. For neutral sentiment, the system performed relatively better, predicting 15 articles as neutral, close to the 12 identified by humans.

These discrepancies suggest that while the system effectively classifies sentiment, it requires fine-tuning, particularly in how it handles subtle or complex language. A potential improvement could involve training the sentiment analysis model with more nuanced and context-aware data, such as sarcasm detection or better differentiation of neutral statements. By improving its handling of such content, the system could provide more accurate insights into inclusivity trends, ensuring that the data used to inform policies and public opinion is both reliable and representative.

4.2. Experiment 2

Another potential blind spot in the BrainBow platform is the accuracy of sentiment analysis when applied to specific topics like race, disability, gender, and hate speech. Sentiment analysis models may struggle with these nuanced and sensitive topics, which can lead to incorrect classifications. Ensuring the system accurately identifies sentiment for each topic is crucial for providing reliable inclusivity metrics.

To test the topic-specific accuracy, we will evaluate how well the system performs sentiment analysis on articles and tweets specifically focused on race, disability, gender, and hate speech. For this experiment, we will use 50 articles or social media posts, evenly distributed across these four topics. Each article or post will be manually annotated by human reviewers to determine its sentiment (positive, neutral, or negative). The system will then run the same set of articles through the sentiment analysis tool to compare the accuracy of its classifications.

We will measure the system's performance for each topic using precision, recall, and F1-score. This comparison will reveal if certain topics are more challenging for the system to analyze correctly, such as articles related to gender issues, which may have more nuanced language.

Figure 8. Figure of experiment 2

The results of the topic-specific sentiment analysis reveal notable differences in the system's accuracy across the four categories. The system demonstrates strong performance in race and hate speech, with high precision, recall, and F1-scores (above 80%). These results suggest that the model can effectively classify sentiment in these areas, likely because discussions on race and hate speech often use explicit or direct language, which is easier for the model to interpret. The consistent performance across these categories indicates that the system is well-suited to capture the public sentiment in more straightforward and polarized discussions.

However, the system struggles with gender and disability, achieving lower precision and recall rates, which results in lower F1-scores. These topics often involve more nuanced, sensitive, or subtle language that the current sentiment analysis model may not fully grasp. Discussions on gender and disability may be framed with complex emotions, implicit biases, or subtleties that the system fails to capture accurately. To improve performance, the system could benefit from more advanced language models or retraining on specialized datasets for these topics, which would help the BrainBow platform provide a more balanced and accurate representation of inclusivity across diverse issues.

5. RELATED WORK

A significant approach for addressing neurodiversity inclusivity involves the creation of a Community of Practice (CoP), as explored by Dettmer and Welton (2023)[11]. This methodology focuses on increasing awareness and providing resources for educators to better support neurodivergent students in higher education. The CoP creates a space for dialogue, knowledge exchange, and resource development, which leads to improved practices. However, while effective at fostering awareness and collaboration, it is limited by the slower rate of institutional change and does not provide real-time data to track inclusivity. In contrast, our project improves upon this by offering real-time data analysis, which allows stakeholders to respond more quickly to inclusivity challenges.

Zsofia et al. (2023) applied time-series data analysis to explore neurodivergent student outcomes in higher education[12]. This methodology uses student performance data over a decade, highlighting key factors that influence the success rates of neurodivergent students, such as scholarships and dormitory placement. While valuable for identifying long-term trends, it does not offer insights into real-time sentiment or public perception, which is crucial for responsive policy changes. Our project builds on this by integrating real-time sentiment analysis, thus providing more immediate feedback on inclusivity efforts.

A third approach, as proposed by Sheikh and Jaiswal (2020), is real-time sentiment analysis using Twitter data[13]. This method effectively classifies tweets based on sentiment, providing real-time insights into public opinion on specific topics. However, it is limited by the quality of the data extracted and the accuracy of machine learning models. Our project extends this methodology by focusing on neurodiversity-related content and combining data from multiple sources to create a more comprehensive inclusivity index.

6. CONCLUSIONS

The BrainBow platform, while effective in providing real-time insights into neurodiversity and inclusivity trends, has several limitations that require further improvement. First, the system's sentiment analysis occasionally struggles with more nuanced or implicit language, especially in topics related to gender and disability. The current model might misclassify or overestimate sentiment polarity, which could lead to skewed or misleading insights. Improving this limitation would require training the system with more diverse and complex datasets, or incorporating advanced NLP models like BERT that can better understand context and subtle expressions [14].
Another limitation is the system's handling of sarcasm, implicit bias, and complex emotions, especially in articles that are not explicitly positive or negative. Additionally, the reliance on specific APIs for data collection introduces potential gaps if certain data sources are unavailable or restricted. A possible improvement would be to expand the range of data sources and ensure backups are in place to maintain consistent real-time analysis [15]. Providing more customization for users, such as personalized filters for specific topics or time ranges, could also enhance user engagement and the utility of the platform.

In conclusion, BrainBow provides a valuable tool for raising awareness and understanding neurodiversity through real-time data analysis. By addressing its current limitations, the platform has the potential to significantly improve the inclusivity metrics it generates, ensuring a more balanced and accurate portrayal of neurodiverse-related discussions in society.

REFERENCES

[1] Das, Sushree, Ranjan Kumar Behera, and Santanu Kumar Rath. "Real-time sentiment analysis of twitter streaming data for stock prediction." Procedia computer science 132 (2018): 956-964.
[2] DeYoung, Jay, et al. "ERASER: A benchmark to evaluate rationalized NLP models." arXiv preprint arXiv:1911.03429 (2019).
[3] Austin, Robert D., and Gary P. Pisano. "Neurodiversity as a competitive advantage." Harvard Business Review 95.3 (2017): 96-103.
[4] Gherardi, Silvia. "Community of practice or practices of a community." The Sage handbook of management learning, education, and development (2009): 514-530.
[5] Ellis, Nick C. "At the interface: Dynamic interactions of explicit and implicit language knowledge." Studies in second language acquisition 27.2 (2005): 305-352.
[6] Lanz, Andreas, Manfred Reichert, and Peter Dadam. "Robust and flexible error handling in the AristaFlow BPM Suite." Information Systems Evolution: CAiSE Forum 2010, Hammamet, Tunisia, June 7-9, 2010, Selected Extended Papers 22. Springer Berlin Heidelberg, 2011.
[7] Al-Hindawi, Fareed Hameed, and Abid Hmood Ali. "A pragmatic study of CNN and BBC news headlines covering the Syrian conflict." Advances in Language and Literary Studies 9.3 (2018): 43-51.
[8] Wilson, Kaitlyn, and Aaron Dallman. "Strategies for promoting neurodiversity and autism acceptance in higher education." Journal of Diversity in Higher Education (2024).
[9] Stephens-Davidowitz, Seth. "The cost of racial animus on a black candidate: Evidence using Google search data." Journal of Public Economics 118 (2014): 26-40.
[10] Gu, Xiaodong, et al. "Deep API learning." Proceedings of the 2016 24th ACM SIGSOFT international symposium on foundations of software engineering. 2016.

[11] Dettmer, Jennie, and Karen Welton. "Increasing neurodiversity awareness through a community of practice." Journal of Learning Development in Higher Education 29 (2023).
[12] Clouder, Lynn, et al. "Neurodiversity in higher education: a narrative synthesis." Higher Education 80.4 (2020): 757-778.
[13] Sheikh, Hisham Ahmed, and Jitendra Jaiswal. "Implementing Sentiment Analysis on Real-Time Twitter Data." J ETIR 7.9 (2020).
[14] Koroteev, Mikhail V. "BERT: a review of applications in natural language processing and understanding." arXiv preprint arXiv:2103.11943 (2021).
[15] Gross, Jürgen H. "Direct analysis in real time—a critical review on DART-MS." Analytical and bioanalytical chemistry 406 (2014): 63-80.