TAMU DATATHON 2023

MARKY CHALLENGE

Introduction:

This project focuses on the critical task of determining user approval or disapproval of AI-generated posts, which include images and textual content. We aim to answer the fundamental question of whether data science and machine learning can effectively gauge user sentiment towards these posts. To achieve this, we leverage computer vision to analyse images and natural language processing to assess text content. The project explores the interplay between visual and textual elements in shaping user sentiments, and we provide insights into our methodology and results. This report outlines our approach, challenges faced, and the innovative aspects of our work, with the goal of contributing to the understanding of AI-generated content and user engagement in the digital sphere.

Team Infinite Matrix:

Shivesh Chowdary Kodali
Pragati Naikare
Jack Wooley
Vinay Chandra

Problem Statement:

Our project aims to predict user approval or disapproval of AI-generated posts, which comprise both textual features and images. This task requires developing a model that can effectively gauge user sentiment towards such posts. This problem has broad implications for content creation, digital marketing, and user engagement, making it a crucial challenge in the age of AI-generated content.

Data Collection and Preparation:

We began with a diverse dataset that presented its own set of challenges. The raw data consisted of a data frame with a mix of categorical variables, text columns, and a column containing links to the associated images.

Image Data Retrieval:

To work with the image data, we executed a script that systematically retrieved each image based on its provided link and stored them in a designated folder. This step allowed us to access the visual component of the posts, a critical element in our analysis.

Text Data Preprocessing:

We applied a series of preprocessing steps to clean and prepare the text data for analysis:

- **Punctuation Removal:** We removed punctuation marks to ensure consistency in the text data and to eliminate any noise that could affect our analysis.
- **Stopword Removal:** Common stopwords, such as "the," "and," and "is," were removed to focus on the most meaningful words and phrases.
- **Contractions Expansion:** We expanded contractions (e.g., "can't" to "cannot") to ensure uniformity and clarity in the text.
- **URL Removal:** Any URLs present in the text were removed to prevent them from interfering with our analysis.
- **Special Character and Number Removal:** Special characters and numerical values were eliminated to create a more text-focused dataset

Finally the data set was split into training and validation data so that we can use validation to predict the best model. But after selecting the best model we combined the train and validation data set to train it again from scratch because of the scarcity of data

Methodology:

(This part explains on how actually we wanted to solve the problem but due to resource constraints we sticked to the method mentioned after this**)**
Our actual approach that we wanted to follow to predicting user approval or disapproval of AI-generated posts involves a fusion of image and text data, leveraging state-of-the-art techniques for interpretability. We present an overview of our methodology:

1. Feature Extraction from Images:

- The image dataset, extracted and stored in a dedicated folder, serves as a rich source of visual information. We converted these images into numerical features using advanced techniques, allowing our model to understand the visual aspects of each post.
- To enhance interpretability, we incorporated Grad-CAM (Gradient-weighted Class Activation Mapping). Grad-CAM enables us to visualize and understand which regions of an image are influential in the model's predictions. This provides valuable insights into the visual cues that drive user sentiment.

2. Text LSTM Model:

- We employed a Text LSTM (Long Short-Term Memory) model to analyze the textual content of the posts. This model can capture sequential patterns in the text, making it suitable for understanding the contextual aspects of the content.
- To ensure transparency and interpretability of the model's decisions, we integrated LIME (Local Interpretable Model-agnostic Explanations). LIME allows us to identify the key features in the text that contribute most to the model's predictions. This enables us to pinpoint the aspects of the text that influence user approval or disapproval.

3. Fusion of Image and Text:

- Our final prediction model combines the features extracted from the images and the output of the Text LSTM. This fusion approach enables us to harness the power of both visual and textual data to make more accurate predictions.

4. Interpretability:

- The use of Grad-CAM for images and LIME for text provides interpretability to our model. We can visually and textually analyze the critical factors that drive the model's decisions. This transparency is crucial in understanding user sentiment towards AI-generated content.

This methodology not only enhances the accuracy of our predictions but also allows us to gain insights into the reasons behind user approval or disapproval of posts, both in terms of visual and textual content.

Our approach to predicting user approval or disapproval of AI-generated posts primarily revolves around text analysis. Due to resource constraints, we opted for a text-based model for this task. Here is an overview of our modified methodology:

1. Text Data Processing:

- We began by thoroughly pre-processing the text data as described earlier, which included steps such as removing punctuation, stopwords, contractions, URLs, special characters, and numbers.

2. Text-Based Model:

- We employed a Text-Based Model, which was designed to analyze the textual content of the posts. This model is well-suited for understanding user sentiment in the absence of image data and served as an effective means to address our project's objectives.

3. Interpretability:

- While we initially planned to integrate Grad-CAM for image interpretability and LIME for text interpretability, resource constraints led us to prioritise the text-based model without visual analysis.

4. Model Training and Evaluation:

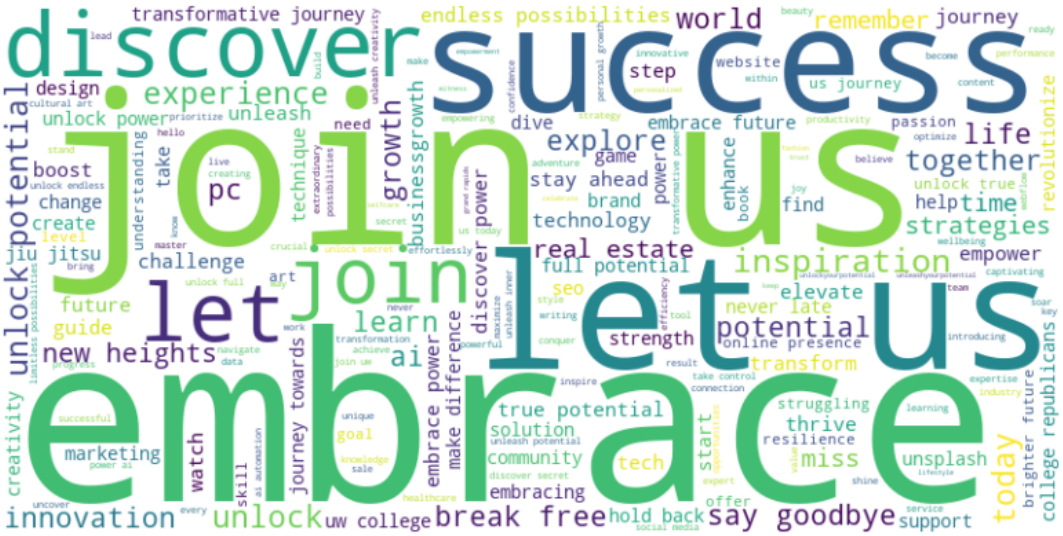
- We trained our text-based model on the pre-processed data, utilising a suitable architecture(LSTM). Evaluation metrics were used to assess the model's performance, providing insights into its ability to predict post approval or disapproval based on text content.

By focusing on text-based analysis, we streamlined our approach to adapt to the available computational resources. This approach allowed us to make meaningful predictions based on textual content while sacrificing the visual analysis aspect of our original methodology.

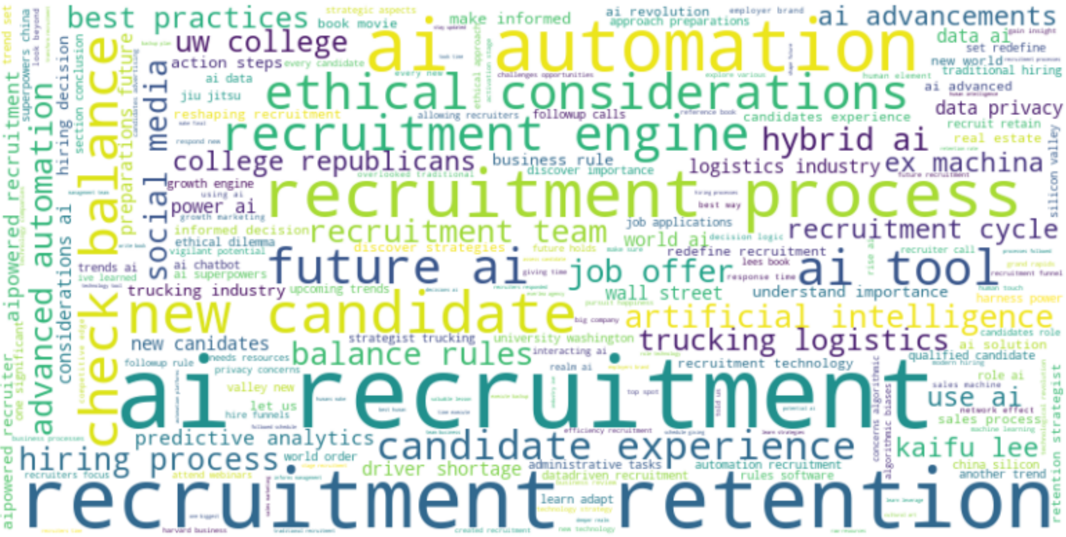
Data Analysis and Results:

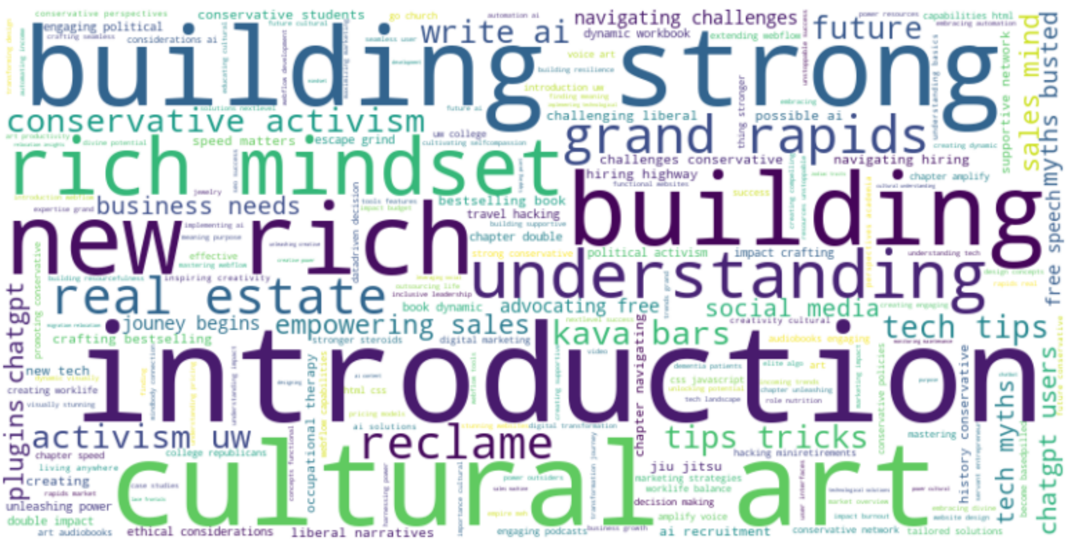
Sample Poster

Word Clouds for columns Caption, Title, Summary, Search Term:




social media
business
ai
strategies
marketing
technology
image
technique
design
success
real estate
website
community
innovation
jiu jitsu
challenge
benefits
tips
webflow
jewelry
life
brand
digital
growth
conservative
college republican
potential
training
support
goal
power
solution
sale
art
financial
importance
tool
creative
building
development
management
performance
engagement
optimization
cultural


ai automation
ai recruitment
recruitment retention
recruitment process
ethical considerations
recruitment engine
new candidate
future ai
ai tool
check balance
advanced automation
candidate experience
hiring process
artificial intelligence
trucking logistics
balance rules
recruitment team
recruitment cycle
hybrid ai
college republicans
best practices
ai advancements
uw college
social media
kaifu lee
job offer
ex machina
predictive analytics
driver shortage
logistics industry
data privacy
use ai


building strong
introduction
cultural art
new rich
building
understanding
rich mindset
grand rapids
conservative activism
real estate
reclame
activism uw
tips tricks
tech tips
kava bars
social media
empowering sales
business needs
future
write ai
navigating challenges
plugins chatgpt
sales mind
myths busted
free speech
chatgpt users
tech myths

Results:

Model	Training Loss	Training Accuracy	Testing Accuracy
Single LSTM 8 Units	0.5273	75.23	73.22
Single LSTM 32 Units	0.5134	76.34	71.06
BERT 7 epochs	0.36	90.24	71.8
BERT 5 epochs	0.42	84.38	73.83
BERT 3 epochs	0.47	79.12	75.25

Challenges Faced:

While pursuing our project, we encountered several noteworthy challenges that influenced our approach and outcomes. The primary challenge revolved around model interpretability, which played a pivotal role in understanding the factors contributing to user approval or disapproval.

1. Interpretability of LSTM Models:

- Interpreting LSTM models was a complex task. To gain insights into the model's decision-making process, we utilised LIME (Local Interpretable Model-agnostic Explanations). This method provided some interpretable results, allowing us to understand how the LSTM-based model arrived at its predictions.

2. Challenges with BERT Models:

- On the other hand, interpreting BERT models proved to be a more demanding endeavour. Despite efforts to employ various interpretability libraries such as eli5, tf-explain, and captum, we encountered multiple errors and obstacles. This posed a significant challenge in extracting insights from the BERT-based model.

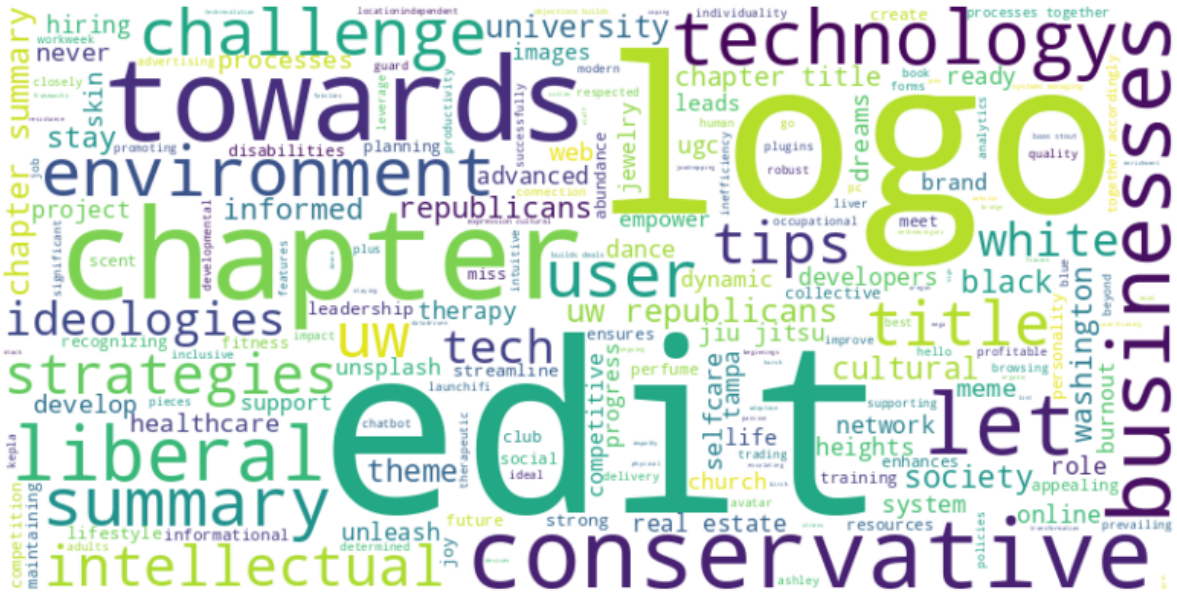
3. Comparative Model Performance:

- Despite the difficulties in interpreting BERT models, they outperformed LSTM models in terms of prediction accuracy by a margin of approximately 2 percent. This underscores the trade-off between model interpretability and predictive power and highlights the importance of addressing both aspects in data science projects.

These challenges were instrumental in shaping our project's trajectory. While LSTM models provided some interpretability using LIME, BERT models excelled in predictive accuracy, even though their interpretability remained elusive.

Here are the most important words while predicting a True labels

Here are the most important words while predicting the False Labels



Future Work:

Our project, while offering valuable insights into the prediction of user approval of AI-generated posts, encountered resource limitations that restricted our ability to fully implement the proposed methodology. As we look ahead, there are several avenues for future work that could enhance the comprehensiveness and depth of our analysis.

1.Comprehensive Fusion of Image and Text Data:

- One promising area for future exploration is the comprehensive fusion of image and text data, as initially planned. With improved computational resources, we can revisit the integration of both modalities, harnessing the combined power of visual and textual cues for more accurate predictions.

2. Advanced Interpretability for BERT Models:

- The challenges faced in interpreting BERT models highlight the need for advancements in model interpretability tools. Future work could focus on developing or adopting enhanced interpretability techniques specifically tailored to complex models like BERT, facilitating a deeper understanding of their decision-making processes.

Conclusion:

In this project, we focused on predicting user approval of AI-generated posts through text analysis, revealing the importance of tone, keywords, and engagement in shaping user sentiment. Despite resource limitations, we learned valuable lessons about the trade-off between model interpretability and predictive power. As we move forward, future work may encompass comprehensive image and text fusion, improved interpretability for complex models like BERT, and larger datasets for enhanced accuracy. In the ever-evolving world of AI-generated content, our project underscores the need for responsible content creation and showcases the potential of data science in shaping digital experiences.

Resources :

https://towardsdatascience.com/interpreting-an-lstm-through-lime-e294e6ed3a03/

https://huggingface.co/docs/transformers/model_doc/bert

https://pypi.org/project/eli5/

https://captum.ai/

https://pypi.org/project/tf-explain/