Research Proposal: Harnessing Data Mining and Text Analytics for Social Media Sentiment Analysis

1. Research Hypothesis and Objective:

Hypothesis:

In the realm of social media sentiment analysis, the integration of sophisticated data mining and text analytics methodologies holds the potential to yield nuanced insights into public opinion dynamics, thereby facilitating more informed decision-making processes for businesses and policymakers alike.

Objectives:

i. **Methodological Advancement:** Developing and implementing an innovative framework that synthesizes cutting-edge data mining techniques and text analysis methodologies tailored specifically for the dynamic and unstructured nature of the social media data.

ii. **Precision and Accuracy:** Enhance the accuracy and granularity of sentiment analysis by leveraging advanced natural language processing(NLP) algorithms to discern subtle nuances in language use and sentiment expression across diverse social media platforms.

iii. **Scalability and Adaptability:** Design a scalable and adaptable sentiment analysis framework capable of accommodating varying data volumes, languages, and contextual factors inherent in social media discourse, ensuring robust performance across different domains and applications.

iv. **Real-time Insights:** Enable real-time monitoring and analysis of social media sentiments trends to facilitate timely response strategies for business and policy makers, empowering them to proactively address emerging issues and capitalize on evolving market sentiments.

v. **Knowledge Dissemination:** Disseminate research findings through peer-review publications, conference presentations, and knowledge-sharing platforms to foster cross-disciplinary collaboration and contribute to the broader academic and professional discourse on social media analytics and sentiment analysis methodologies.

vi. **Validation and Benchmarking:** Conduct rigorous validation and benchmarking exercises to assess the efficacy and reliability of the proposed framework against established benchmarks and ground truth datasets, ensuring the integrity and generalizability of the findings.

2. Background:

In contemporary society, social media platforms have emerged as pervasive channels for communication, information dissemination, and public discourse. The unprecedented growth of social media usage has engendered as a digital ecosystem characterized by the continuous generation of vast volumes of user-generated content, ranging from text-based posts and comments to multimedia content such as images and videos.

Sentiment analysis, also known as opinion mining, entails the computational analysis of textual data to discern the subjective sentiment or emotional tone expressed within the text. The primary objective of sentiment analysis is to categorize text into sentiment categories, such as positive, negative, or neutral, based on the underlying emotional polarity conveyed by the text. By analyzing sentiment patterns and trends in social media data, businesses, policymakers, and researchers can gain valuable insights into public opinion, consumer preferences, brand perception, and socio-political dynamics.

Historically, sentiment analysis on social media has predominantly relied on rule-based approaches, lexicon-based methods, and simple machine learning algorithms for sentiment classification. These traditional techniques often struggled to capture the nuances and complexities of human language, particularly in the context of informal communication styles, slang, sarcasm, and cultural references prevalent on social media platforms. Hover, recent advancements in data mining and text analytics have paved the way for more sophisticated and nuanced approaches to sentiment analysis, capable of overcoming many of the limitations inherent in traditional methodologies.

Key advancements driving the evolution of sentiment analysis on social media include:

- **Deep Learning**: Deep learning techniques, such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), and transformer-based models like BERT and GPT, have demonstrated remarkable performance in capturing semantic and contextual information from text data, enabling more accurate and contextually-aware sentiment analysis.

- **Multimodal Analysis**: With the proliferation of multimedia content on social media, researchers are increasingly exploring multimodal approaches that integrate textual, visual, and audio features for sentiment analysis, allowing for a more holistic understanding of sentiment expressed across different modalities.

- **Cross-Lingual Sentiment Analysis**: The globalization of social media has necessitated the development of cross-lingual sentiment analysis techniques capable of analyzing sentiment in multiple languages, facilitating broader reach and applicability of sentiment analysis frameworks across diverse linguistic contexts.

Past research endeavours have laid a foundational groundwork for sentiment analysis on social media, driven by the imperative to understand and interpret the collective sentiment of online communities. These efforts have contributed seminal methodologies and techniques, spanning from lexicon-based approaches to machine learning-based classification algorithms, aimed at discerning the polarity and emotional valence embedded within textual data. Moreover, advancements in deep learning architectures and neural network models have further propelled the frontier of sentiment analysis, enabling more nuanced and contextually-aware sentiment interpretation.

Despite these advancements, the landscape of social media sentiment analysis remains rife with challenges and complexities. The dynamic and heterogeneous nature of social media discourse, characterized by slang, sarcasm, and cultural nuances, poses inherent obstacles to accurate sentiment interpretation. Furthermore, the sheer volume and velocity of social media data necessitate scalable and efficient analysis frameworks capable of processing and extracting actionable insights in real-time.

3. Importance and contribution to knowledge:

In the contemporary digital landscape, the importance of harnessing data mining and text analytics for social media sentiment analysis cannot be overstated. This research endeavour holds significant potential to contribute to various domains, including business intelligence, public opinion monitoring, policymaking, and academic research. The following points elucidate the importance and potential contributions of this research:

i. **Enhancing Business Decision-Making:**
   - By leveraging advanced data mining and text analytics techniques, businesses can gain deeper insights into customer sentiment, preferences, and brand perception on social media platforms.
   - Analyzing social media sentiment enables businesses to identify emerging trends, anticipate customer needs, and tailor marketing strategies and product offerings accordingly, thereby enhancing customer engagement and satisfaction.
   - Moreover, sentiment analysis can aid in reputation management, crisis response, and competitive analysis, allowing businesses to mitigate risks and capitalize on opportunities in real-time.

ii. **Informing Public Policy and Governance:**
   - Policymakers and government agencies can benefit from social media sentiment analysis by monitoring public opinion, identifying societal concerns, and gauging the effectiveness of policies and initiatives.
   - Insights derived from sentiment analysis can inform evidence-based policymaking, crisis management strategies, and public communication campaigns, fostering greater transparency, accountability, and responsiveness in governance.

iii. **Addressing Societal Challenges:**
   - Social media sentiment analysis can play a pivotal role in addressing pressing societal challenges, such as public health crises, environmental issues, and social justice movements.
   - By analyzing sentiment trends and discourse on social media, researchers and advocacy groups can identify emerging issues, mobilize public support, and advocate for policy changes and social reforms.

iv. **Advancing Academic Research**:
   - The proposed research contributes to the broader academic discourse on data mining, text analytics, and sentiment analysis methodologies.
   - By exploring novel techniques for sentiment analysis on social media data, the research extends the frontier of knowledge in computational linguistics, machine learning, and artificial intelligence.
   - Moreover, the research generates empirical insights into the dynamics of public sentiment in digital environments, offering valuable datasets and benchmarks for future research endeavours in related fields.

v. **Fostering Interdisciplinary Collaboration:**
   - The interdisciplinary nature of this research fosters collaboration between academia, industry, and government sectors, facilitating knowledge exchange and technology transfer.
   - Collaborative partnerships enable the translation of research outcomes into real-world applications, driving innovation, and socioeconomic impact in diverse domains.

In summary, harnessing data mining and text analytics for social media sentiment analysis has profound implications for businesses, policymakers, researchers, and society at large. By unlocking the latent insights embedded within social media data, this research endeavour contributes to informed decision-making, public engagement, and knowledge advancement, ultimately shaping the future of digital communication and societal discourse.

4. Pilot study:

In the realm of entertainment industry analysis through data mining and text analytics, the scope encapsulates a comprehensive examination of audience sentiment across various entertainment domains, including movies, television shows, music releases, celebrity news, and events. This analysis extends to encompassing multiple social media platforms, such as Twitter, Facebook, Instagram, YouTube, and specialized entertainment forums, where discourse on entertainment content proliferates. The overarching aim is to leverage advanced analytical techniques to discern patterns in audience sentiment, discern trends, and ascertain the impact of promotional efforts and content releases within the entertainment landscape.

The objectives of this endeavour are manifold: Firstly, to delve into audience sentiment dynamics, thereby facilitating a deeper understanding of audience preferences and reactions towards specific entertainment offerings. Secondly, to evaluate the efficacy of content strategies by assessing audience sentiment and engagement metrics derived from social media data. Additionally, the pursuit aims to identify emerging trends, influential voices, and potential reputation risks within the entertainment sphere. Moreover, this research endeavour seeks to inform strategic decision-making processes of entertainment stakeholders, aiding in the optimization of marketing strategies, content development, and audience engagement initiatives. Ultimately, by delineating the scope and objectives in this manner, the study aims to equip industry stakeholders with actionable insights gleaned from data-driven analysis, fostering informed decision-making and fostering sustained growth within the entertainment industry landscape.

Sentiment analysis, a prominent task in Natural Language Processing (NLP), plays a pivotal role in understanding the public opinion, consumer behaviour, and social trends. Leveraging state-of-the-art machine learning techniques, this pilot study aims to explore sentiment analysis methodologies using PyTorch and the Hugging Face Transformers library. By harnessing the por of deep learning models such as BERT (Bidirectional Encoder Representation from Transformers), this research endeavour seeks to analyze sentiment in textual data with enhanced accuracy and efficiency.

The primary objective of this pilot study is to evaluate the effectiveness of PyTorch-based implementations of transformer models, specifically BERT, for sentiment analysis tasks. By employing transfer learning techniques and fine-turning pre-trained models on sentiment analysis datasets, the study aims to assess the performance of these models in capturing nuanced sentiment patterns across different domains and languages. Furthermore, the research seeks to investigate the impact of various factors such as dataset size, model architecture, and training strategies on sentiment analysis performance.

The methodology for this plot study encompasses four key steps, firstly, data acquisition involves in curating a diverse array of sentiment analysis datasets covering various domains, languages and sentiment categories. By leveraging publicly available datasets and domain-specific corpora, my aim to ensure comprehensive coverage of sentiment expression, enabling robust model training and evaluation.

Following data acquisition, the model selection and pre-processing phase involve the careful selection of appropriate transformer-based models from the Hugging Face

Transformers library, such as BERT, tailored for sentiment analysis tasks. This step also includes pre-processing textual data through tokenization, padding, and special token insertion to prepare input sequences for subsequent model training and evaluation stages.

Subsequently, the methodology proceeds to model training and fine-tuning, where transfer learning techniques are employed to fine-tune pre-trained transformer models on the acquired sentiment analysis datasets. Leveraging Pytorch’s flexible framework, custom training pipelines, loss functions, and optimization strategies are defined to cater to specific sentiment analysis objectives.

Finally, the evaluation and performance metrics stage assess the effectiveness of trained models using standard sentiment analysis metrics like accuracy, precision, recall, and f1-score.Through comprenhsive experimentation, the study analyses the impact of hyper parameters, training data size, and domain specific on model performance, providing insights into optimal model configurations for sentiment analysis tasks across diverse domains and languages.

The successful implementation of sentiment analysis using BERT has significant implications for various applications, including brand monitoring, market research, and social media analytics. By accurately classifying sentiments expressed in social media posts, organizations can gain valuable insights into customer opinions, preferences, and trends.

Moving forward, future research could explore advanced techniques for sentiment analysis, such as incorporating domain-specific embedding’s or ensemble models to further improve performance. Additionally, extending the study to analyze the real-time social media data streams and integrating sentiment analysis into the decision-making processes could enhance its practical utility.

5. Programme and Methodology:

Data Acquisition and Pre-processing:

The first phase of the study involves data acquisition and pre-processing to prepare the dataset for sentiment analysis. A diverse set of sentiment analysis datasets encompassing various domains, languages, and sentiment categories is curated. The dataset, obtained from a CSV file named “sentimentdataset.csv” is loaded into a Pandas data frame for further processing. The dataset comprises columns such as “Text” containing textual data and “Sentiment” indicating sentiment labels. Initial exploratory data analysis (EDA) is conducted to understand the dataset’s characteristics, including shape, summary statistics, missing values, and class distribution.

```
df= pd.read_csv("sentimentdataset.csv")
df.shape
df.describe().loc[['min', '50%', 'mean', 'max',
'std']].T.style.background_gradient(axis=1)
df.info()
df.isna().sum()
df_columns= df.columns
for col in df.columns:
  print(col)
df.head()
df.duplicated().any()
```

Model Selection and Pre-processing:

In the subsequent phase, appropriate transformer-based models from the Hugging Face Transformers library are selected for sentiment analysis tasks. Specifically, the BERT (Bidirectional Encoder Representations from Transformers) model with the 'Bert-base-cased' configuration is chosen as the primary model architecture. The BERT tokenizer is utilized for tokenization, padding, and special token insertion to prepare input sequences for model training and evaluation. Additionally, text length distribution analysis is performed to gain insights into the dataset's textual characteristics and inform preprocessing decisions.

```
PRE_TRAINED_MODEL_NAME = 'Bert-base-cased'
tokenizer = BertTokenizer.from_pretrained(PRE_TRAINED_MODEL_NAME)
sample_txt = 'Enjoying a beautiful day at the park!'
tokens = tokenizer.tokenize(sample_txt)
token_ids = tokenizer.convert_tokens_to_ids(tokens)
print(f' Sentence: {sample_txt}')
print(f' Tokens: {tokens}')
print(f'Token IDs: {token_ids}')
```

Model Training and Evaluation:

The core of the methodology involves model training and evaluation using PyTorch. A custom PyTorch dataset class, GPReviewDataset, is implemented to pre-process the textual data and create data loaders for efficient model training. The dataset is split into training, validation, and test sets using the train_test_split function from Scikit-learn. The training data loader, validation data loader, and test data loader are created with batch size set to 8 for efficient processing.

A sentiment classification model, SentimentClassifier, is defined as a subclass of nn.Module in PyTorch. The model architecture consists of a BERT model for feature extraction, follow by a dropout layer and a linear layer for classification. The model is initialized with the pre-trained BERT ights ('bert-base-cased') and adapted for sentiment analysis tasks. Model training is performed using the AdamW optimizer and a linear learning rate scheduler.

```
encoding = tokenizer.encode_plus(
sample_txt,
max_length=32,
add_special_tokens=True, # Add '[CLS]' and '[SEP]'
return_token_type_ids=False,
pad_to_max_length=True,
return_attention_mask=True,
truncation=True,
return_tensors='pt',  # Return PyTorch tensors
```

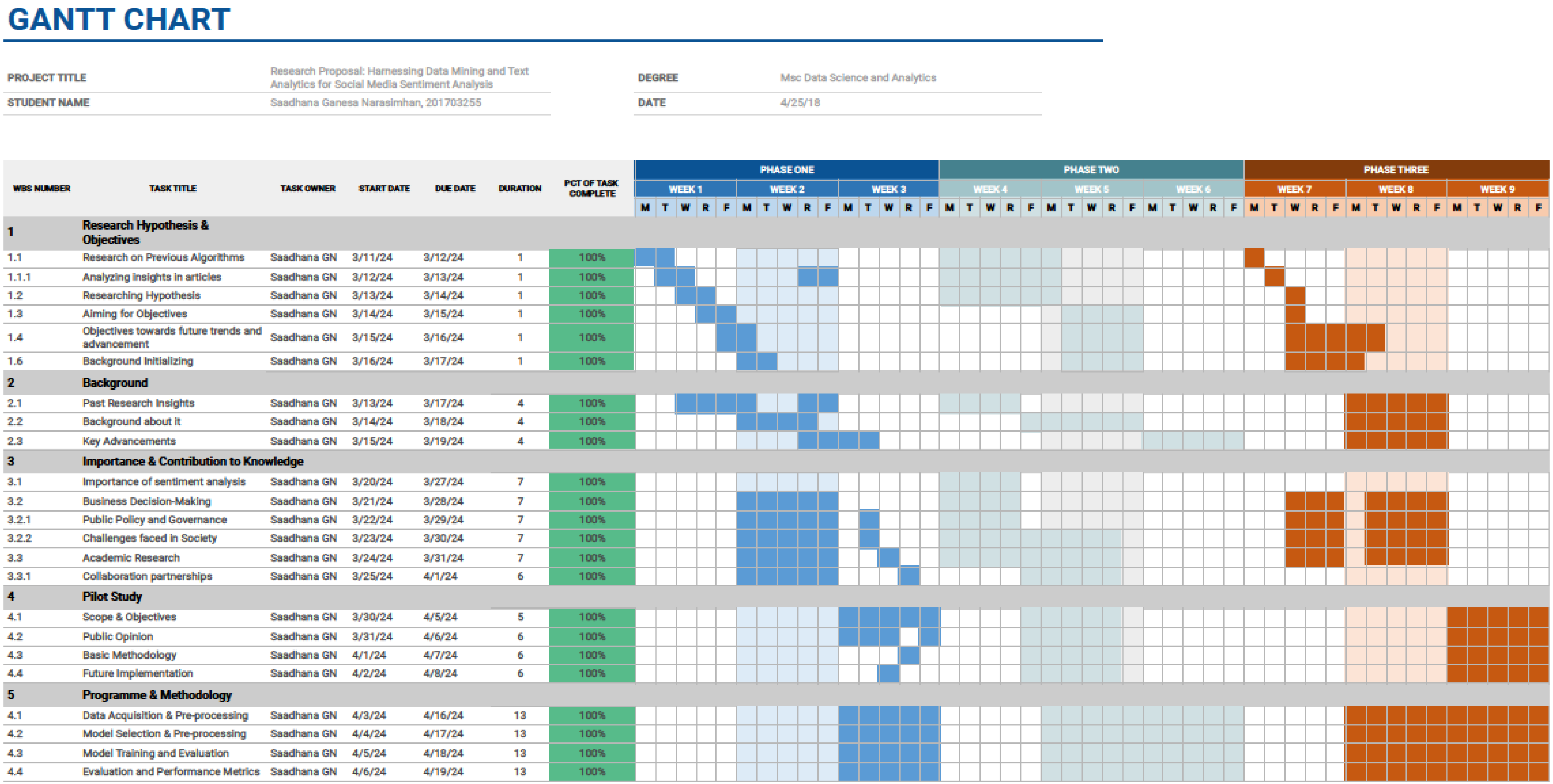
```
)
encoding.keys()
token_lens = []
for txt in df.Text:
  #tokenizing the text
      tokens = tokenizer.encode(txt, max_length=512)
      token_lens.append(len(tokens))
sns.distplot(token_lens)
plt.xlim([0, 256]);
plt.xlabel('Token count');
class GPReviewDataset(Dataset):
  def __init__(self, reviews, targets, tokenizer, max_len):
    self.reviews = reviews
    self.targets = targets
    self.tokenizer = tokenizer
    self.max_len = max_len
  def __len__(self):
    return len(self.reviews)
  def __getitem__(self, item):
    review = str(self.reviews[item])
    target = self.targets[item]
    encoding = self.tokenizer.encode_plus(
      review,
      add_special_tokens=True,
      max_length=self.max_len,
      return_token_type_ids=False,
      pad_to_max_length=True,
      return_attention_mask=True,
      return_tensors='pt',
    )
    return {
      'review_text': review,
      'input_ids': encoding['input_ids'].flatten(),
      'attention_mask': encoding['attention_mask'].flatten(),
      'targets': torch.tensor(target, dtype=torch.long)
    }
df_train, df_test = train_test_split(df, test_size=0.30, shuffle=True)
df_val, df_test = train_test_split(df_test, test_size=0.50,shuffle=True)
df_train.shape, df_val.shape, df_test.shape
class GPReviewDataset(Dataset):

  def __init__(self, reviews, targets, tokenizer, max_len):
    self.reviews = reviews
    self.targets = targets
    self.tokenizer = tokenizer
    self.max_len = max_len
  def __len__(self):
    return len(self.reviews)
  def __getitem__(self, item):
    review = str(self.reviews[item])
    target = self.targets[item]
    encoding = self.tokenizer.encode_plus(
      review,
      add_special_tokens=True,
      max_length=self.max_len,
      return_token_type_ids=False,
      pad_to_max_length=True,
      return_attention_mask=True,
      return_tensors='pt',
    )
    return {
      'review_text': review,
      'input_ids': encoding['input_ids'].flatten(),
      'attention_mask': encoding['attention_mask'].flatten(),
      'targets': torch.tensor(target, dtype=torch.long)
    }
df_train, df_test = train_test_split(df, test_size=0.30, shuffle=True)
df_val, df_test = train_test_split(df_test, test_size=0.50,shuffle=True)
```

```
df_train.shape, df_val.shape, df_test.shape
def create_data_loader(df, tokenizer, max_len, batch_size):
  ds = GPReviewDataset(
    reviews=df.Text.to_numpy(),
    targets=df.Sentiment.to_numpy(),
    tokenizer=tokenizer,
    max_len=max_len,
  )
  return DataLoader(
    ds,
    batch_size=batch_size,
    num_workers=4,
    shuffle=True
  )
BATCH_SIZE = 8
train_data_loader = create_data_loader(df_train, tokenizer, MAX_LEN, BATCH_SIZE)
val_data_loader = create_data_loader(df_val, tokenizer, MAX_LEN, BATCH_SIZE)
test_data_loader = create_data_loader(df_test, tokenizer, MAX_LEN, BATCH_SIZE)
data = next(iter(train_data_loader))
data.keys()
```

Evaluation and Performance Metrics:

The final phase of the methodology involves model evaluation and performance metrics calculation. The trained model is evaluated on the validation and test datasets using standard sentiment analysis metrics, including accuracy, precision, recall, and F1-score. Classification reports and confusion matrices are generated to assess the model's performance across different sentiment categories. The evaluation results provide insights into the model's effectiveness in capturing sentiment patterns and its generalization capabilities.

6. Work plan Diagram:



GANTT CHART

PROJECT TITLE	Research Proposal: Harnessing Data Mining and Text Analytics for Social Media Sentiment Analysis	DEGREE	Msc Data Science and Analytics
STUDENT NAME	Saadhana Ganesa Narasimhan, 201703255	DATE	4/25/18

WBS NUMBER	TASK TITLE	TASK OWNER	START DATE	DUE DATE	DURATION	PCT OF TASK COMPLETE
1	**Research Hypothesis & Objectives**					
1.1	Research on Previous Algorithms	Saadhana GN	3/11/24	3/12/24	1	100%
1.1.1	Analyzing insights in articles	Saadhana GN	3/12/24	3/13/24	1	100%
1.2	Researching Hypothesis	Saadhana GN	3/13/24	3/14/24	1	100%
1.3	Aiming for Objectives	Saadhana GN	3/14/24	3/15/24	1	100%
1.4	Objectives towards future trends and advancement	Saadhana GN	3/15/24	3/16/24	1	100%
1.6	Background Initializing	Saadhana GN	3/16/24	3/17/24	1	100%
2	**Background**					
2.1	Past Research Insights	Saadhana GN	3/13/24	3/17/24	4	100%
2.2	Background about it	Saadhana GN	3/14/24	3/18/24	4	100%
2.3	Key Advancements	Saadhana GN	3/15/24	3/19/24	4	100%
3	**Importance & Contribution to Knowledge**					
3.1	Importance of sentiment analysis	Saadhana GN	3/20/24	3/27/24	7	100%
3.2	Business Decision-Making	Saadhana GN	3/21/24	3/28/24	7	100%
3.2.1	Public Policy and Governance	Saadhana GN	3/22/24	3/29/24	7	100%
3.2.2	Challenges faced in Society	Saadhana GN	3/23/24	3/30/24	7	100%
3.3	Academic Research	Saadhana GN	3/24/24	3/31/24	7	100%
3.3.1	Collaboration partnerships	Saadhana GN	3/25/24	4/1/24	6	100%
4	**Pilot Study**					
4.1	Scope & Objectives	Saadhana GN	3/30/24	4/5/24	5	100%
4.2	Public Opinion	Saadhana GN	3/31/24	4/6/24	6	100%
4.3	Basic Methodology	Saadhana GN	4/1/24	4/7/24	6	100%
4.4	Future Implementation	Saadhana GN	4/2/24	4/8/24	6	100%
5	**Programme & Methodology**					
4.1	Data Acquisition & Pre-processing	Saadhana GN	4/3/24	4/16/24	13	100%
4.2	Model Selection & Pre-processing	Saadhana GN	4/4/24	4/17/24	13	100%
4.3	Model Training and Evaluation	Saadhana GN	4/5/24	4/18/24	13	100%
4.4	Evaluation and Performance Metrics	Saadhana GN	4/6/24	4/19/24	13	100%

7. References:

1. Jacob, D. (2018, Oct 11). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. https://huggingface.co/papers/1810.04805
2. Moez Ali (Mar 2023). NLTK Sentiment Analysis Tutorial for Beginners. https://www.datacamp.com/tutorial/text-analytics-beginners-nltk
3. Marius, M. (2021, Jan 13). Sentiment Analysis: First Steps With Python's NLTK Library. https://realpython.com/python-nltk-sentiment-analysis/
4. Federico, P. (2022, Feb 2). Getting Started with Sentiment Analysis using Python. https://huggingface.co/blog/sentiment-analysis-python#2-how-to-use-pre-trained-sentiment-analysis-models-with-python
5. Sunil, S. (2023, July 6). Using ChatGPT for Sentiment Analysis: A Beginner's Guide. https://blog.gramener.com/using-chatgpt-for-sentiment-analysis/#:~:text=ChatGPT's%20ability%20to%20understand%20natural,ideal%20tool%20for%20sentiment%20analysis.
6. Rami, J. (2023, Jun 18). Harnessing the Power of Python and Machine Learning: Sentiment Analysis of Social Media Data — Rami Jaloudi. https://medium.com/@rjaloudi/harnessing-the-power-of-python-and-machine-learning-sentiment-analysis-of-social-media-data-78b91fb2247b
7. Abeer, M. (2023, Aug 22). Best Social Media Sentiment Analysis Tools: Unveiling Insights With Data. https://www.aimtechnologies.co/best-social-media-sentiment-analysis-tools-unveiling-insights-with-data/
8. Granis, C. (2016). Social Media Analytics in the entertainment industry. Information and Communication Systems. https://repository.ihu.edu.gr/xmlui/bitstream/handle/11544/15227/c.granis_icts_06-04-2017.pdf?sequence=1
9. FasterCapital. (2024, Mar 9). Sentiment analysis: Harnessing Emotional Insights with Social Media Analytics. https://fastercapital.com/content/Sentiment-analysis--Harnessing-Emotional-Insights-with-Social-Media-Analytics.html#:~:text=Sentiment%20analysis%20uses%20natural%20Language,positive%2C%20negative%2C%20or%20neutral.
10. Ines, R. (2019, Oct 28). 15 of The Best Sentiment Analysis Tools. https://monkeylearn.com/blog/sentiment-analysis-tools/
11. Kashish, P. (2024, Jan). Social Media Sentiments Analysis Dataset. Kaggle https://www.kaggle.com/datasets/kashishparmar02/social-media-sentiments-analysis-dataset

8. Appendix:

In developing the report, employed a combination of data mining and text analytics tools to facilitate various stages of the research process. Here's a breakdown of our use of these tools:

Data Collection and Pre-processing:

- For the small pilot study, utilized Python libraries such as pandas and scikit-learn for data manipulation and preprocessing tasks. These libraries allow us to efficiently handle datasets, perform data cleaning, and prepare the data for analysis.
- Employed the Transformers libraries, specifically the BertTokenizer and BertModel classes, for tokenization and encoding of text data. This pre-processing step was crucial for inputting textual data into the BERT model for sentiment analysis.

Model Training and Evaluation:

- The pilot study involved training and fine-tuning the BERT model for sentiment analysis tasks. Leveraged PyTorch, a deep learning framework, for implementing the model architecture, defining custom training pipelines, and conducting model evaluation.

- Evaluation metrics such as accuracy, precision, recall, and F1-score re calculated using scikit-learn's classification_report and confusion_matrix functions. These metrics provided insights into the performance of the trained model on the validation dataset.

Literature Review and Background Information:

- To gather background information and relevant research studies, and utilized academic search engines like Google Scholar. Queries such as "Sentiment Analysis using BERT" and "Transformer-based models for Social Media Sentiment Analysis" re employed to retrieve scholarly articles, conference papers, and research papers related to our topic.

- Additionally, used ChatGPT to generate summaries, draft sections of the report, and refine our understanding of complex concepts. For instance, interacted with ChatGPT by asking questions like "Can you summarize the benefits of using BERT for sentiment analysis?" or "What are the limitations of existing sentiment analysis techniques?"

Visualization and Reporting:

- Matplotlib and Seaborn libraries re utilized for data visualization tasks, including plotting histograms, pair plots, and distribution plots. These visualizations helped us explore the characteristics of the dataset and communicate key findings effectively.

- Finally, Microsoft Word was used for drafting and formatting the report. The report was structured according to the guidelines provided, with sections covering research hypothesis, background, methodology, results, and conclusions.

Overall, the use of these data mining and text analytics tools enabled us to conduct a comprehensive pilot study on sentiment analysis using BERT, gather relevant literature, analyze data, train machine learning models, and report our findings accurately and effectively.