Dream Interpretation through Machine Learning and NLP: A Research Study

Tanvee Proag (2422814)

University of Mauritius

Abstract. Dreams are a part of human beings' subconscious mind. The subconscious mind is the brain's ability to perceive and process information. This paper is a study on the interpretation of dreams through the use of natural language processing and machine learning techniques. The aim is to have a better understanding of people's dreams, find patterns in dreams and observe correlation between dreams and mental health. The research question that will guide this study is *How can AI models be trained to interpret dream narratives?* It will do so by trying to identify emotional as well as symbolic patterns and the connection between dream content and the psychological states. This research adds to both AI advancements and psychological studies to understand the subconscious mind as it can help with mental health self-assessment. The results show an average topic diversity of above 0.9 which suggests that the topics are well-separated and have distinct vocabularies. This shows that the model is performing well. The results of this research could have several practical applications such as assisting therapists in understanding patients' subconscious thoughts and helping individuals in their self-reflection and in their personal growth.

Keywords: Dreams, Dream Interpretation, Mental Health, Psychotherapy, Artificial Intelligence, Machine Learning, NLP

---

1 Introduction

Human beings' ability to dream has been a source of intrigue for many centuries. The Babylonians and Sumerians attempted to unravel this mystery by categorizing dreams into *good* and *bad*, attributing them to their mythical gods and demons respectively. In ancient Greece, figures such as Antiphon believed that dreams are the gateway to prophecies and could serve as guidance. Meanwhile, the Chinese held that dreams involved the soul journeying outside the body during sleep [Wiki 2019].

In the modern era, dreams remain fascinating. Sigmund Freud's *The Interpretation of Dreams* posited that dreams reflect unconscious desires and unresolved conflicts [Freud 1932]. Carl Jung's *Structure and Dynamics of the Psyche* expanded on Freud's work, focusing on the collective unconscious and personal symbolism in dreams [Jung 1981]. Today, dream interpretation continues to grow with diverse psychological and cultural approaches. It serves as an aid for increased personal awareness and also a therapeutic technique in psychotherapy.

Dreams are considered windows into the subconscious mind because they reflect unconscious thoughts, emotions and memories. Both historical and contemporary theories agree on the correlation between the nature of dreams and their connection to the subconscious. Today, neuroscientific research supports the notion that dreams are linked to the subconscious. During REM sleep, brain regions associated with emotion, memory and self-processing are active which suggests that dreams help integrate and consolidate emotional experiences into long-term memory. [Scalabrini et al. 2021] Dreams often reflect unconscious memories from early childhood and attachment-related experiences which are stored implicitly in memory. Thus, they can serve as a means of emotional self-state expression, helping to process and reorganize emotional experiences. Moreover, they facilitate communication between the conscious and subconscious mind, allowing for findings on thoughts and emotions. [Hinckley 2023]

Artificial Intelligence has the ability to revolutionize the field of dream interpretation by tapping into the subconscious mind and retrieving patterns. This integration of AI with traditional dream analysis techniques looks promising as it can improve scientists' understanding of dreams thus providing new tools dedicated to personal growth and self-awareness.

AI models have the ability to quickly analyze vast amounts of data, identifying recurring themes, emotions and symbols in dreams that may not be immediately apparent to human interpreters. This significantly reduces personal biases and offers a clear understanding of dream symbolism and significance. There exists AI tools that can process and interpret dreams quickly, making dream analysis more accessible and user-friendly. Platforms like *Dreamology* and *Temenos* provide free or low-cost access to AI-driven dream interpretation. [DemoWP 2024, Temenosdreams 2025]

2 Literature Review

For data collection, there exist multiple data banks online.

Domhoff and Bulkeley established a dreambank (dreambank.net) comprising 20,000 dream reports and a spreadsheet program [Bulkeley et al. 2010]. These help in calculating the dream content ratios that were derived from the major Hall-Van de Castle categories. Even though the dreambank.net spreadsheet proved helpful in the study of dream content, there was a steep learning curve as researchers had to learn a detailed scoring system that was associated with the Hall-Van de Castle categories. Then, they had to code all dream reports prior to the calculation of normative ratios in the spreadsheet manually.

A study on dream content analysis used Dreamboard for its rich dataset of dream content [Das et al. 2022]. In the Dreamboard internet forum that started in 2012, users record and track their dreams online over time. After narrating their post in text format, they then themselves categorize or make comments on their dreams. The entries they post comprise a mixture of standardized fields and open text fields to capture their personal dream narratives and themes. For confidentiality, identification codes are assigned once the data goes into the database. In 2017, the website made publicly available a database of almost 38,000 dreams of the last three years. The limitation of this study was that it only focused on dreams posted in English.

Reddit's 44,213 dream reports that consist of 217 topics and grouped into 22 larger themes can be used [Das et al, 2022]. It is believed to be the most extensive collection of dream topics currently - A crowd-sourced dataset of dream self-reports from the r/Dreams community on Reddit. Unlike traditional lab studies, dream experiences shared on r/Dreams are reported voluntarily and spontaneously, enabling the collection of a large set of dream reports and conducting an ecological study. Over 44K dream reports from more than 34K Reddit users were collected over the past five years.

Another study that followed the previous one used Reddit dream reports and justified that it was the best option because unlike prior dream research that relied heavily on traditional laboratory, survey and diary methods or dreams that were recorded by waking up people from REM sleep and asking for their input so as to increase dream recall, Reddit dream reports are not influenced by laboratory setting [Das et al. 2024]. Reddit dream reports unlike survey responses do not suffer from memory distortion since the dreamers write about their dreams immediately instead of waiting for a period of time. Reddit dream reports unlike diary studies are not as ecological with minimal retrospective bias but they are not limited by small sample size from the burden of participation. Another benefit of using Reddit is that it is a social media platform that is popular which means that its usage is growing in numbers and it might continue to be a source of dream content for analysis.

The first to carry out a study about dream analysis was Hall and Van de Castle who developed a classification system based on the scientific criteria that is objective, generalised and reproducible [Hall et al. 1966]. Topics can be validated by comparing them to the Hall and Van de Castle scale. The usual technique to gain insights from dream topics used to be by sifting through many reports manually. The Hall and Van de Castle method offers a supervised approach for content analysis by using predetermined categories. However, these are usually biased towards existing knowledge on dreams. In reality, dreams are weird in nature and they consist of impossible or unlikely events that are not common in the daily life of a real person. This means that dreams may not be found in the predetermined categories that already exist. That is why current approaches are likely to miss some aspects of dreams [McNamara et al. 2019]. Unsupervised methods may prove better suited for dream analysis than supervised methods for a better understanding of dreams to address this gap in knowledge.

3 Methodology

The data is collected, preprocessed and cleaned before the modeling using machine learning techniques and sentiment analysis. Finally, the system shall be evaluated using a set of metrics.

***Data collection* -** The public dream database Reddit r/Dreams posts is used as it is extensive and keeps on growing. One advantage worth noting is that since the dreamers write down their dreams on the social forum, they can do so immediately after waking up which avoids bias caused by forgetting or memorizing not vividly enough. On top of that, open-ended questionnaires are sent out to participants willing to note down their dreams. This might be beneficial for contextual symbolism as people from different cultures and backgrounds might assign different meanings to different symbols. Stratified sampling is used - some people from different countries provide for the raw data as the website is online and international. This ethnic representation decreases bias. In addition, some Mauritians are asked for their dream narratives. This survey is done through Google Forms and the responders are made aware that their responses are anonymous as their emails are not collected. [See Appendix for the Google Forms questionnaire]. Since it is a qualitative study, the questionnaire contains open-ended questions that allow the responders to describe their dreams in great detail. The questions are obligatory and are required to be filled before submitting.

***Pre-processing* -** The text is cleaned by removing unnecessary elements such as stop words (e.g., 'and', 'the'), punctuation marks and special characters to simplify the text and to reduce computational complexity. Then, the text is split into smaller units i.e. tokens, which are typically words or phrases that form the basis to analyze further. The NLTK (Natural Language Toolkit) library is used to identify and label common dream symbols like *water*, *animals*, *death* and *flying*. It is important that these symbols are recognized now to link the dream content to their psychological meaning later. VADER is used for sentiment analysis - it classifies the emotional tone of the dream narratives. This helps to identify the emotions associated with the dream content.

***Feature Extraction* -** The raw text is transformed into a structured format that machine learning models can process. Bag of Words (BoW) is applied to represent dream content numerically by counting word frequencies and then focusing on significant terms compared to common ones. NRC Emotion Lexicon is a pre-trained lexicon that is used to assign emotional labels to words. This allows the model to detect the hidden feelings within the dreams. For semantic similarity detection, BERT is used to detect symbolic patterns in the text. For example, words like *falling* could be associated with anxiety or fear. This improves the model's ability to interpret.

***Modeling* -** Machine learning and deep learning techniques can be used to model the system for dream analysis. The sentiment analysis model uses VADER to classify the overall emotional tone of the dreams. For example, *joy*, *sadness*, *fear* etc. The LSTM (Long Short-Term Memory) model analyzes sequences in the dream narratives to capture patterns and shifts in the dream sequences.

A fine-tuned BERT model that is pre-trained is used on the dream dataset to detect meaning related to symbols in the texts. For example, it may recognize that *falling* in a dream symbolizes anxiety. The data is classified into twenty different categories: *Surreal or Symbolic Dreams*, *Taboo or Uncomfortable Relationship Dreams*, *Dreams About Meaning & Metaphysics, Everyday or Repetitive Dreams*, *Death or Violent Dreams*, *Time or Memory Distortion Dreams*, *Anxiety or Observation-Based Dreams*, *Social or Family Anxiety Dreams*, *Coming-of-Age or Embarrassment Dreams*, *Medical Dreams*, *Transformation or Identity Dreams*, *Narrative or Adventure Dreams*, *Animal & Nature Symbolism Dreams*, *Cultural or Mythical Dreams*, *Shared or Paranormal Experiences*, *Existential or Reality-Bending Dreams*, *Confusion or Desire-Based Dreams*, *Fear of Death or Dissolution*, *Absurd or Creative Dreams*, *Substance-Influenced or Chaotic Dreams.*

4 Data Analysis & Discussions

The model categorizes dream narratives into twenty different emotional categories. Common dream elements are identified by the system.

Latent Dirichlet Allocation (LDA) (from the SKLEARN library) is configured to map the topics into their categories. LDA represents each document as a mixture of topics where each topic is a probability distribution over words. To convert the text data into numerical features suitable for LDA, TF-IDF vectorization is used. This is done to make sure that the importance of words across documents is weighted based on their frequency and rarity. After training the model, the top 10 most representative words are extracted for each topic. Then, each topic is labeled based on a qualitative assessment of the top words. For example, a topic containing words like *death*, *dream*, *died*, *funeral*, *spirit* is interpreted as "*Death or Violent Dreams*" and another topic with terms like *school*, *forgot*, *late*, *test* is categorized as "*Everyday or Repetitive Dreams*". This labeling process is subjective. It converts raw statistical output to meaningful human interpretations. Finally, each dream entry is assigned the topic that has the highest probability score. This is the dominant theme of that post. This creates a categorical variable that can be used for analysis and visualization. For topic distribution, a bar chart is plotted. [See Appendix] It reveals an even spread across several categories. However, certain topics appeared more frequently, suggesting that some themes might be more prevalent in people's dream narratives.

Two more metrics are used to evaluate the validity and reliability of the study - topic diversity and topic confidence. The topic diversity score shows how distinct the top N words across topics are. The result averaged across multiple runs is 0.92. This proves that there was minimal word repetition among topics, that is, each topic covers unique semantic ground. It implies that the model successfully extracted diverse themes with minimal overlap. This indicates a good model quality. The topic confidence score represents how dominant a particular topic is within each document. This was assessed by taking the maximum topic probability for each document. A histogram plot of this data showed that most documents had high confidence scores (most of them above 0.60) which shows that the model could confidently assign a dominant theme to most dreams. [See Appendix]

To visualize the relationships between documents based on their topic distributions, t-distributed Stochastic Neighbor Embedding (t-SNE) is applied. Its use is to reduce the topic space to two dimensions. Each point in the scatterplot represents a single dream post which is colored by its assigned topic. t-SNE is kind of a clustering algorithm which helps visualize latent patterns and possible clusters. The plot showed distinguishable clusters of dream narratives which shows that the topic modeling is effective in identifying distinct dream types. [See Appendix] There is a few overlap which is expected as dream narratives are complex and multi-themed.

5 Conclusion & Recommendations

The analysis shows that it is feasible to apply topic modeling to dream narratives through natural language processing and machine learning. The diversity score is high which shows that the clusters of themes are clear and the model has uncovered the interpreted topic from the dream narratives input. The plots show that dreams can be interpreted.

There were some challenges during the implementation phase. Different people assign different meanings to the same dream symbols. Therefore, dream interpretation is highly subjective. So, it is reasonable that the AI model's predictions and the human interpretations are inconsistent. Symbolic representations in dreams are ambiguous in nature. One symbol, like *water*, might have multiple interpretations depending on the context of the dream and on the individual's personal experiences. Also, there is a lack of large annotated dream datasets. This limits the model's ability to make generalizations across a large number of dream narratives. In addition, integrating mental health predictions based on dream content without clinical diagnosis can create misdiagnosis. Therefore, it must be approached with caution. The study can be repeated with more input. The topics are assumed to be independent for simplicity. However, for complex narratives, it may not be true. A more advanced model, like NMF, can be used together with more raw data.

Even though these challenges exist, the potential benefits of this research outweigh the risks. The development of an AI-based dream interpretation system could add to emotional well-being of people and help with mental health self-assessment [Jenni et al. 2024]. The system is automated. This makes dream interpretation accessible and consistent for people who need it.

Recommendation - The Delphi method which is a systematic forecasting method that relies on a panel of experts can be used to refine the interpretation of topics. The experts could be a dozen psychologists and psychiatrists who can validate thematic categorizations through a census. First, the professionals could be consulted and given the findings - a report that contains the topics extracted, the representative keywords, sample dreams and visualizations collected. Round one would be an initial feedback where the experts are asked to label and describe each topic in their own words, request feedback and make suggestions. Then, agreement and disagreement on the topic agreements could be sought and the contradictions could be highlighted. New topics could be proposed. For round two, the summaries are to be sent back to the experts who are asked to re-evaluate their initial opinions after seeing the group's insights. They can revise their interpretations and agree with the others. For round three, there should be a consensus on the agreed topic labels, psychological interpretations and suggestions to improve the model's categories. The Delphi method would ensure that the model can prove valid and reliable in clinical use.

References

1. Arxiv.org. (2022). Dream Content Discovery from Reddit with an Unsupervised Mixed-Method Approach. [online] Available at: `https://arxiv.org/html/2307.04167` [Accessed 28 Feb. 2025].
2. Bulkeley, K., & Domhoff, G. W. (2010). Detecting meaning in dream reports: An extension of a word search approach. Dreaming, 20, 77-95.
3. Das, A., Sanja Scepanovic, Aiello, L.M., Mallett, R., Barrett, D. and Quercia, D. (2024). Dream Content Discovery from Social Media Using AI. Research Square (Research Square). [online] doi:`https://doi.org/10.21203/rs.3.rs-4549419/v1`.
4. DemoWP. (2024). Dreamology Insights — Unlock the Hidden Meanings Your Dreams. [online] Available at: `https://dreamologyinsights.com/` [Accessed 1 Mar. 2025].
5. Freud, S. (1932). The interpretation of dreams. London: Allen & Unwin.
6. Hall, C. S., & Van De Castle, R. L. (1966). The content analysis of dreams. Appleton-Century-Crofts.
7. Heidt, A. (2025). Why Are Recurring Dreams Usually Nightmares? [online] Scientific American. Available at: `https://www.scientificamerican.com/article/why-are-recurring-dreams-usually-nightmares/` [Accessed 1 Mar. 2025].
8. Hinckley, C. (n.d.). Looking into the window of your subconscious mind. [online] The Spellbinder. Available at: `https://thespellbinder.net/8501/features/looking-into-the-window-of-your-subconscious-mind/` [Accessed 27 Feb. 2025].
9. Jenni, P. and Roesler, C. (2024). A New Way of Analysing Dreams on Its Profoundest Level: The Development of Motif Analysis and Phase Model (MAP) as an Extension of Structural Dream Analysis (SDA). Behavioral Sciences, [online] 14(8), pp.658–658. doi:`https://doi.org/10.3390/bs14080658`.
10. Jung, C.G. (1981). The structure and dynamics of the psyche. Princeton, N.J.: Princeton University Press.
11. McNamara, P., Duffy-Deno, K., Marsh, T. and Marsh, T.J. (2019). Dream content analysis using Artificial Intelligence. International Journal of Dream Research, [online] pp.42–52. doi:`https://doi.org/10.11588/ijodr.2019.1.48744`
12. Priory (2023). Mental Health Statistics. [online] Priory. Available at: `https://www.priorygroup.com/mental-health/mental-health-statistics` [Accessed 28 Feb. 2025].

13. Scalabrini, A., Mucci, C. and Esposito, R. (2021). Dreaming the unrepressed unconscious and beyond: repression vs dissociation in the oneiric functioning of severe patients. Research in Psychotherapy: Psychopathology, Process and Outcome, 24(2). doi:https://doi.org/10.4081/ripppo.2021.545.
14. Temenosdream.com. (2025). Temenos Dreams. [online] Available at: https://www.temenosdream.com/ [Accessed 1 Mar. 2025].
15. Wikipedia Contributors (2019). Dream. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Dream [Accessed 27 Feb. 2025].
16. World Health Organization (2024). Mental Health. [online] World Health Organization. Available at: https://www.who.int/health-topics/mental-health#tab=tab_1 [Accessed 28 Feb. 2025].

6 Appendix

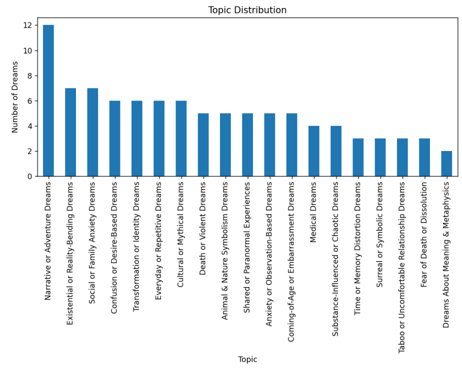


Fig. 1. Topic Distribution

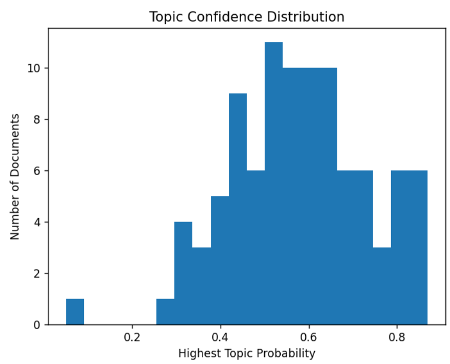


Fig. 2. Topic Confidence

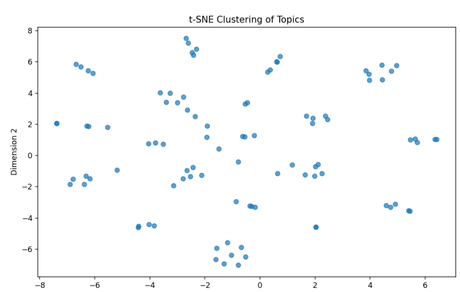
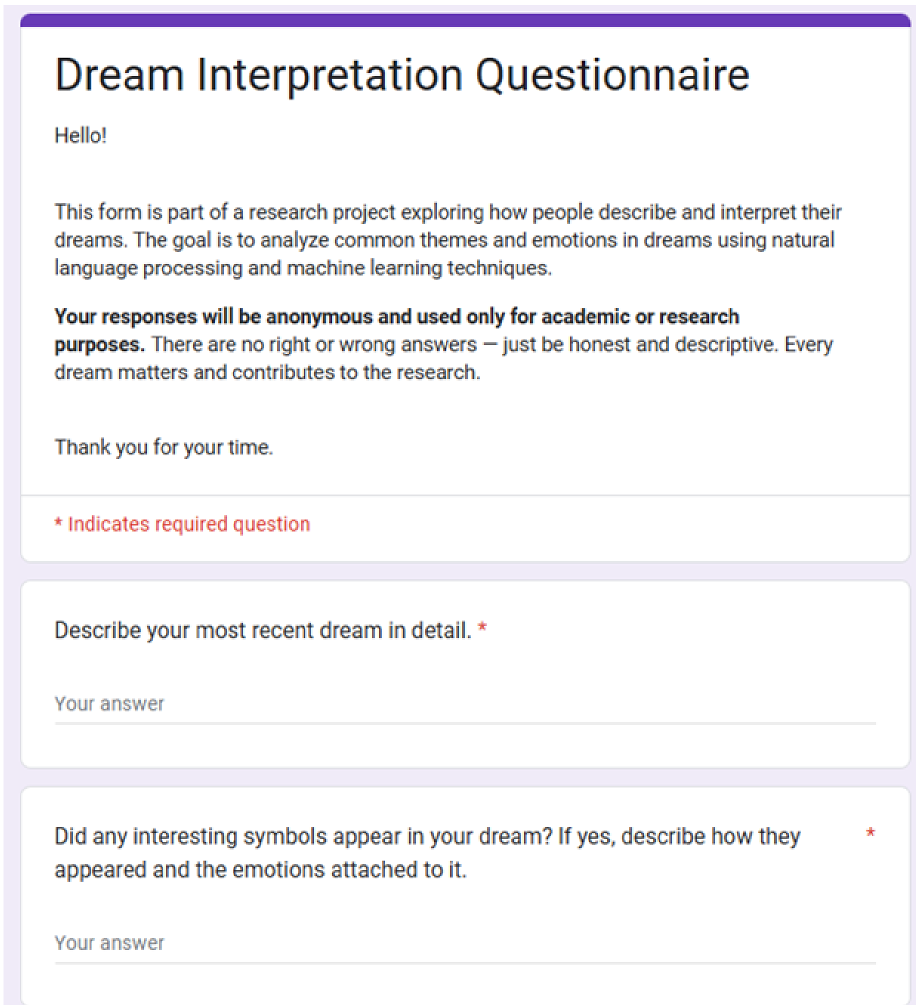


Fig. 3. t-SNE plot



Dream Interpretation Questionnaire

Hello!

This form is part of a research project exploring how people describe and interpret their dreams. The goal is to analyze common themes and emotions in dreams using natural language processing and machine learning techniques.

Your responses will be anonymous and used only for academic or research purposes. There are no right or wrong answers — just be honest and descriptive. Every dream matters and contributes to the research.

Thank you for your time.

* Indicates required question

Describe your most recent dream in detail. *

Your answer

Did any interesting symbols appear in your dream? If yes, describe how they appeared and the emotions attached to it. *

Your answer

Fig. 4. Google Forms Questionnaire