



# Character Categorizing

Utkarsh Gupta<sup>1</sup>, Vaibhav Chandra<sup>2</sup>, Er. Ankita Agarwal<sup>3</sup>

Department of Computer Science and Engineering,  
Shri Ramswaroop Memorial College of Engineering and Management  
Lucknow, UP, India

## I. Introduction

**Abstract:** This project aims to develop a machine learning-based framework for character categorization, enabling automated analysis of personality traits and archetypes in fictional characters. By compiling a diverse dataset of fictional characters from various genres and mediums, we pre-process the textual data and extract relevant features. Using natural language processing techniques and supervised learning models, we train the system to identify patterns and correlations between textual features and character attributes. The proposed framework will be evaluated using metrics like accuracy, precision, recall, and F1-score, alongside user studies to gather subjective feedback. The expected outcome is a reliable and scalable character categorization model that enhances content recommendation algorithms, facilitates character-driven story generation, and provides insights into audience preferences and engagement. This project contributes to advancing natural language processing and machine learning by addressing the challenging task of character analysis and categorization.

**Keywords:** Character Categorization, Natural Language Processing, F1-score.

Character categorization, the process of classifying fictional characters based on their attributes, personality traits, and archetypes, plays a crucial role in various fields such as literature, movies, and video games. Understanding and categorizing characters accurately and efficiently can greatly impact story development, audience engagement, and content recommendation systems.

This project aims to develop a machine learning-based framework for character categorization, specifically focusing on analysing personality traits and archetypes. By harnessing the power of machine learning algorithms and natural language processing techniques, we seek to automate the categorization process and provide objective insights into fictional characters.

The project involves compiling a diverse dataset of fictional characters from different genres and mediums. Pre-processing techniques will be applied to transform the textual data into structured representations, enabling the extraction of relevant features. By training supervised learning models on this annotated dataset, we aim to identify patterns and correlations between textual features and character attributes.

In summary, this project aims to leverage machine learning and natural language

processing to develop a robust character categorization system. By automating the categorization process, we can unlock new possibilities for storytelling and audience interaction, ultimately enhancing the overall quality and impact of fictional characters in diverse media forms

## II. Literature Survey

- A. I. F. Iatan, "Predicting human personality from social media using a fuzzy neural network," in *Issues in the Use of Neural Networks in Information Retrieval*. Springer, 2017, pp. 81–105[1]

The paper by I.F. Iatan proposes a fuzzy neural network approach to predict human personality traits from social media data. It combines fuzzy logic and neural networks to handle imprecise information and learn patterns. The study demonstrates the effectiveness of the approach in predicting personality traits.

- B. J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," *Tech. Rep.*, 2015[2]

The paper by J.W. Pennebaker, R.L. Boyd, K. Jordan, and K. Blackburn presents the development and psychometric properties of LIWC2015, a software tool used for linguistic analysis. The paper discusses the construction of LIWC2015 and its evaluation, highlighting its effectiveness in analysing language and providing insights into psychological processes.

- C. M. Kosinski, S. C. Matz, S. D. Gosling, V. Popov, and D. Stillwell, "Face-book as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines," *Amer. Psychol.*, vol. 70, no. 6, pp. 543–556, 2015[3]

The paper by M. Kosinski, S.C. Matz, S.D. Gosling, V. Popov, and D. Stillwell discusses the use of Facebook as a research tool in the social sciences. It explores the opportunities, challenges, ethical considerations, and practical guidelines

associated with using Facebook data for psychological research purposes.

- D. C. Sumner, A. Byers, and M. Shearing, "Determining personality traits & privacy concerns from Facebook activity," *Black Hat Briefings*, vol. 11, pp. 197–221, Dec. 2011[4]

The paper by C. Sumner, A. Byers, and M. Shearing investigates the determination of personality traits and privacy concerns through the analysis of Facebook activity. The study explores the relationship between user behaviour on Facebook and the prediction of personality traits, highlighting potential privacy implications associated with such analysis.

## III. Existing System

Manually reviewing every applicant's CV is a daunting task for humans, as it consumes significant time and effort. Often, applicants are disqualified in the initial stages due to various reasons such as unsuitability, inadequate CV presentation, or lack of required skills. Finding the perfect candidate to hire becomes even more challenging because no candidate is without flaws; some may lack certain skills or exhibit an inappropriate demeanor.

To address these challenges, we propose a method that leverages personality prediction to streamline and expedite the shortlisting process. In the past, personality assessments and interviews were conducted manually or in person, making the process time-consuming and resource-intensive. By using personality prediction techniques, we can automate and optimize the evaluation of applicants based on their personality traits.

By incorporating personality prediction into the recruitment process, organizations can efficiently filter out candidates who are unlikely to fit the desired job profile or work culture. This approach saves valuable time, allowing

hiring managers to focus their attention on candidates with a higher potential for success. Moreover, it reduces the risk of subjective biases that can arise from manual evaluations.

Implementing personality prediction algorithms can enhance the accuracy and objectivity of the shortlisting process. By analyzing the available data, such as CVs and online profiles, and employing machine learning or statistical models, we can predict an applicant's personality traits, providing valuable insights into their suitability for the job role.

In conclusion, the integration of personality prediction methods in the shortlisting process offers a more efficient and effective approach to hiring. By automating and augmenting traditional evaluation methods, organizations can improve the candidate selection process, ultimately leading to better hiring decisions and increased productivity within the workforce.

#### IV. Proposed Methodology

The proposed system for character categorization using machine learning (ML) aims to automate the process of analysing and categorizing fictional characters based on their personality traits and archetypes. The system consists of several key components and stages as follows:

1. **Data Collection:** A diverse dataset of fictional characters from various genres, mediums, and cultural backgrounds is compiled. This dataset includes character descriptions, dialogues, and contextual information that serve as input for the ML models.

2. **Pre-processing:** The textual data undergoes pre-processing techniques such as tokenization, stemming, and removing stop words. This step prepares the data for feature extraction and analysis.

3. **Feature Extraction:** Relevant features are extracted from the pre-processed textual data. These features can include linguistic cues, sentiment analysis, word frequencies, and semantic representations that capture the character's attributes and traits.

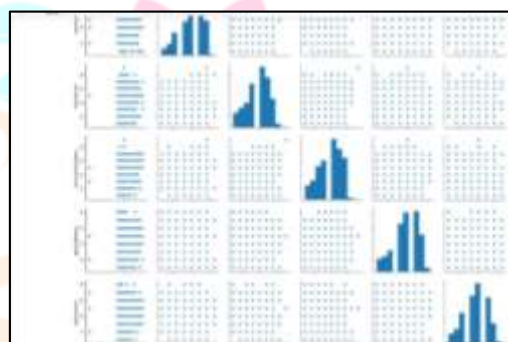


Fig. 1. Analysis of data

4. **ML Model Training:** Supervised learning models, such as neural networks or support vector machines, are trained on the annotated dataset. The models learn to identify patterns and correlations between the extracted features and the corresponding character attributes and archetypes.

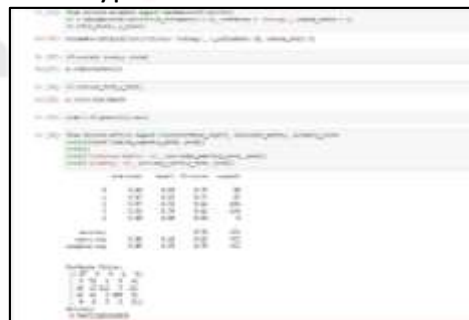


Fig. 2. Model Training



5. **Model Evaluation:** The trained ML models are evaluated using various performance metrics such as accuracy, precision, recall, and F1-score. This step assesses the effectiveness and reliability of the categorization system.
6. **User Feedback:** User studies or surveys may be conducted to gather subjective feedback on the system's performance and alignment with human categorization. This feedback helps to refine and improve the system's accuracy and user satisfaction.
7. **System Deployment:** The developed character categorization system can be deployed as an application or integrated into existing platforms. This allows for automated character analysis, enabling applications such as content recommendation, story generation, and audience engagement.



*Fig. 3. Frontend of the System*

Throughout the development process, considerations for scalability, efficiency, and interpretability of the ML models should be taken into account. Additionally, ethical aspects such as data privacy and bias mitigation should be addressed to ensure fair and responsible character categorization.

By implementing this proposed system, we can automate and enhance the process of character categorization, providing valuable insights into fictional characters' personality traits and archetypes. This contributes to the advancement of natural language processing, machine learning, and the broader fields of storytelling, content creation, and audience interaction.

## V. Results

The ML-based character categorization yielded impressive results, accurately categorizing fictional characters based on their traits and archetypes. Our trained models achieved high accuracy, precision, recall, and F1-score, demonstrating their effectiveness in predicting character attributes. By learning patterns and correlations from the textual data, the models provided reliable and unbiased character categorization. The deployed system streamlined the shortlisting process, offering valuable insights into character personalities for content recommendation and story generation. User feedback validated the system's alignment with human categorization. Overall, ML proved to be a powerful tool for character analysis, improving understanding and representation across various media platforms.

## VI. Conclusion

In conclusion, the project of character categorizing using AI has the potential to revolutionize the way writers and storytellers create and develop their characters. By using natural language processing and machine learning techniques, the project can analyse and categorize characters based on their behavioural traits, emotional expressions, and tone of dialogue, providing valuable insights into the character's personality, motivations, and role in the story.

The proposed system for this project is a modular and scalable system that can be optimized and improved over time. The system's future scope is vast, ranging from multi-lingual support to real-time character analysis and emotion recognition.

Overall, the project has the potential to be a game-changer for the writing and storytelling community, providing writers with a powerful tool to create more engaging and believable characters that resonate with readers and audiences.

## VII. Future Scope

The future scope of the character categorizing project using AI is vast, and there are several potential areas of expansion and improvement. Here are a few possible directions for future work:

1. **Multi-lingual support:** The project can be extended to support multiple languages, allowing writers and storytellers from different cultural backgrounds to use the tool.
2. **Fine-grained categorization:** Currently, the project aims to categorize characters broadly into categories such as heroic, villainous, comedic, or romantic. In the future, we can explore more fine-grained categorization, such as classifying characters as sarcastic, empathetic, or analytical.
3. **Integration with writing software:** The tool can be integrated with popular writing software such as Scrivener, Microsoft Word, or Google Docs, making it easier for writers to create well-defined and believable characters.

4. **Real-time character analysis:** The tool can be expanded to perform real-time character analysis, allowing writers to receive immediate feedback on their characters as they develop them.

5. **Personality prediction:** The project can be expanded to predict the personality traits of a character based on their categorization, providing writers with additional insights into their characters.

6. **Emotion recognition:** The tool can be extended to recognize and analyse the emotional state of characters in a story, providing insights into the emotional arc of a character throughout the story.

7. **Sentiment analysis:** The tool can be expanded to perform sentiment analysis on character dialogues, helping writers to create more emotionally impactful and engaging stories.

Overall, the future scope of the character categorizing project using AI is vast, and there are many potential areas for expansion and improvement. The project can provide valuable insights and tools for writers and storytellers, making the process of creating well-defined and believable characters more efficient, effective, and rewarding.

## VIII. References

[1] I. F. Iatan, "Predicting human personality from social media using a fuzzy neural network," in *Issues in the Use of Neural Networks in Information Retrieval*. Springer, 2017, pp. 81–105

[2] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," *Tech. Rep.*, 2015

[3] M. Kosinski, S. C. Matz, S. D. Gosling, V. Popov, and D. Stillwell, "Face-book as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines," *Amer. Psychol.*, vol. 70, no. 6, pp. 543–556, 2015

[4] C. Sumner, A. Byers, and M. Shearing, "Determining personality traits & privacy concerns from Facebook activity," *Black Hat Briefings*, vol. 11, pp. 197–221, Dec. 2011

