



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 14, Issue 3, March 2025**

# **AI-Powered Text Summarization and Question & Answering System**

**N Bala Yesu, A Lakshmi Niharika, G Harsha Pavan Teja, B Balu, J Abhiram**

Assistant Professor, Department of CSE (AI&ML), Vasireddy Venkatadri Institute of Technology, Guntur, India

U.G. Student, Department of CSE (AI&ML), Vasireddy Venkatadri Institute of Technology, Guntur, India

U.G. Student, Department of CSE (AI&ML), Vasireddy Venkatadri Institute of Technology, Guntur, India

U.G. Student, Department of CSE (AI&ML), Vasireddy Venkatadri Institute of Technology, Guntur, India

U.G. Student, Department of CSE (AI&ML), Vasireddy Venkatadri Institute of Technology, Guntur, India

**ABSTRACT:** The rapid growth of digital text data during our current period requires advanced processing techniques to find meaningful results. An AI-based text processing platform brings Natural Language Processing (NLP) together with deep learning techniques and hybrid machine learning methods to improve text summarization and question answering and speech synthesis abilities and timeline alignment functions and feedback testing analysis. The system executes summarization techniques through TextRank and utilizes pre-trained transformer models BART, T5 and GPT together with extractive and abstractive methods. The system utilizes BERT and ALBERT together with a question answering feature and offers voice support through both speech to-text and text-to-speech conversion capabilities. The core innovation brings a timeline-based content alignment system that preserves logical sequence in data series such as procedural steps and chronological events. The hybrid sentiment analysis framework which uses ALBERT along with BERT and CatBoost detects sentiment orientation and recovers precious user feedback insights. The system bases its accessibility and security features on OTP authentication methods together with profile management functionality. Python and Django serve as backend processing engines in combination with HTML, CSS and JavaScript for implementing frontend development of the proposed solution. The developed system operates across multiple roles to fulfill the requirements of academic investigation as well as content production and enterprise feedback evaluation and assistive system usage.

**KEYWORDS:** Text Summarization, NLP, Deep Learning, Sentiment Analysis, Transformer Models, Speech Synthesis, Question Answering, User Authentication

## **I. INTRODUCTION**

The expansive text data available today represents options for both chance and difficulties in our current digital environment. Efficient text processing techniques face an expanding need due to rapid growth in online content along with academic papers and customer reviews as well as other textual resources. Manual analysis of extended datasets requires an impractical amount of time thereby becoming inefficient. The requirement demands AI solutions at an advanced level to automate operations such as text summarization in addition to question answering while also handling speech synthesis and feedback analysis.

The proposed text processing system employs latest Natural Language Processing (NLP) and deep learning together with hybrid machine learning models to improve how users handle textual information. Users benefit from multiple features in this system where they can produce text summaries and obtain exact question responses while sentiment analysis identifies feedback emotions and the system ensures logical structure in sequential content. The text summarization and question answering performance of the system reaches high accuracy levels through its implementation of BART, T5, GPT, BERT and ALBERT transformer models. The system offers text and speech conversion abilities which extend its usefulness to users with different needs.

This system brings an innovative timeline-based content alignment capability that maintains both historical events together with procedural steps and sequential information in logical structures. A sentiment analysis model

uniting ALBERT with BERT and CatBoost delivers enriched analytical capabilities for user feedback evaluation which allows businesses together with researchers to achieve better audience sentiment understanding. AP College Guide serves students by uniting machine learning algorithms with database organization and NLP processes to link careers with student possibilities. Through this platform students can find complete career guidance and college selection resources in one centralized location thus improving efficiency and using data-based decisions.

OTP-based authentication together with user profile management features exist in the system to ensure security and ease of use. The system applies Django alongside Python for processing backend operations and uses HTML along with CSS and JavaScript to build its frontend structure. The system supports database management through SQLite together with MySQL thus providing adaptability for different application needs.

The system delivers specific advantages to students along with researchers and content developers as well as enterprises together with assistive technology applications. This automated text processing system achieves better efficiency and user accessibility together with better engagement through its ability to manage complex tasks independently.

## II. LITERATURE SURVEY

The quick advancement of Natural Language Processing (NLP) together with deep learning techniques radically changed the way users process text for analysis and summary creation. Different research approaches developed during several years have contributed to the improvement of text summarization alongside question answering technology and sentiment analysis and speech synthesis methods. The combination of extractive and abstractive text processing methods through hybrid models has created efficient systems that handle context-based content analysis.

The development of text summarization moves from traditional statistical models toward deep learning-based advanced solutions. The original text summarization tools included Latent Semantic Analysis (LSA) and Term Frequency-Inverse Document Frequency (TF-IDF) which were commonly used for extractive summarization during that period [1], [2]. The approaches failed to detect semantic connections between words at the same time they produced many redundant summary results. The TextRank algorithm established graph-based models to determine essential sentences within texts through their prominence in the document [3].

The advent of deep learning along with transformer-based models like BART, T5 and GPT improved abstractive summarization quality at a high level [4]. Through their utilization of self-attention mechanisms these models acquire contextual understanding which allows them to create summaries that resemble human writing. Through the Hugging Face library businesses can implement transformer-trained models to create accessible and effective text summarization solutions [5].

The question answering domain underwent a transformation when BERT and ALBERT transformer models proved able to obtain exact answers from extensive text databases [6], [7]. Visor consistent understanding processes enable the models to deliver precise responses based on their contextual analysis. Users today can easily access AI systems through voice commands because of combined speech-to-text and text-to-speech technologies [8].

The research field observes Timeline-based content alignment as a fundamental element for handling structured text systems. Sequential information needs proper logical organization to achieve both coherence and readability for events and procedure steps. The predictive summarization techniques developed recently empower NLP models to deduce missing text segments thus improving the output content quality as reported in [9].

The analysis of sentiment remains vital for performing feedback analysis while obtaining customer insights. Sentiment classification traditionally relied on Naive Bayes alongside Support Vector Machines (SVM) until software limitations restricted effective contextual evaluation of texts [10]. ALBERT BERT and CatBoost models represent recent sentiment detection advances through their combination of semantic meaning processing with gradient boosting algorithms [11].

Organizations use OTP-based verification because it functions as a strong protection system against unauthorized access to maintain user security. Research documents establish OTP based authentication as a security solution which strengthens online platforms by minimizing credential-based attack risks [12].

The new AI text summarization system along with its feedback analysis component merges these technical developments to create a functional text processing system that users can easily use. Hybrid machine learning models and deep learning alongside NLP strengthen the system to deliver precise summary outputs and faster question response while generating valuable sentiment results. The system maintains its wide range of applications because it includes speech synthesis and secure authentication features which aid numerous business sectors including education, enterprise feedback management and assistive technologies.

### III. METHODOLOGY

The proposed AI-powered text summarization and feedback analysis system integrates Natural Language Processing (NLP), deep learning, and hybrid machine learning models to provide efficient text processing. This section details the architecture, techniques, and algorithms employed in the system.

#### A. System Architecture

The architecture of the system is composed of multiple layers, including data preprocessing, text analysis, model execution, and result generation. The core components interact seamlessly to perform summarization, question answering, speech synthesis, content alignment, and feedback sentiment analysis.

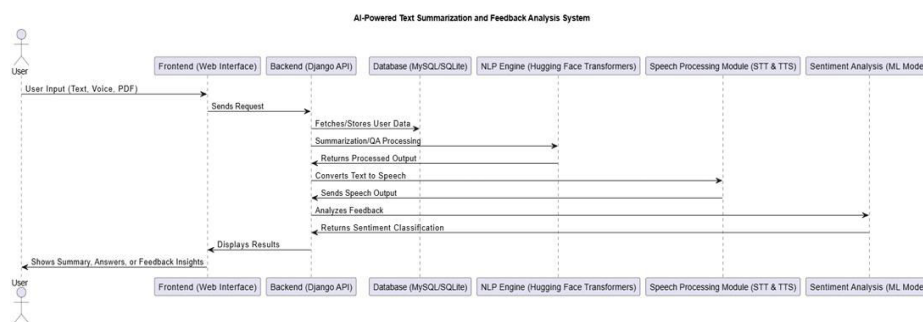


Fig. 1. Overall System Architecture

#### B. Text Summarization

Text summarization is implemented using both extractive and abstractive approaches. The extractive summarization method utilizes the TextRank algorithm, which is based on a graph-based ranking approach. The importance of a sentence is determined by analyzing its connections with other sentences. Mathematically, the sentence ranking is computed as follows:

$$S(V_i) = (1 - d) + d \sum_{V_j \in In(V_i)} \frac{S(V_j)}{|Out(V_j)|}$$

where:

- $S(V_i)$  is the score of sentence  $V_i$ ,
- $d$  is the damping factor (typically set to 0.85),
- $In(V_i)$  represents sentences that link to  $V_i$ ,
- $Out(V_j)$  represents sentences to which  $V_j$  links.

For abstractive summarization, transformer-based models such as BART, T5, and GPT are utilized. These models generate human-like summaries by understanding contextual dependencies within the text.

**C. Question Answering with Speech Support**

The question-answering module employs BERT and AL BERT transformer models to extract precise answers from a given text. Speech recognition is incorporated using Speech Recognition for converting spoken queries into text. The response is synthesized using gTTS or PyTTSX3 for vocal output.

**D. Time Line Based Content Alignment**

The system maintains logical order in textual data, ensuring that events, processes, and instructional steps follow a coherent sequence. This is particularly useful for historical content and procedural documentation. Predictive summarization is used to infer missing segments of the text.

**E. Feedback Analysis and Sentiment Detection**

Sentiment analysis is performed using a hybrid machine learning model combining ALBERT, BERT, and CatBoost. These models work together to classify feedback into positive, neutral, and negative sentiments. CatBoost enhances classification accuracy through gradient boosting.

$$P(y|X) = \frac{e^{f(X)}}{1 + e^{f(X)}}$$

Where:

- P(y|X) is the probability of sentiment y given input X,
- f(X) is the learned function mapping input text to sentiment scores.

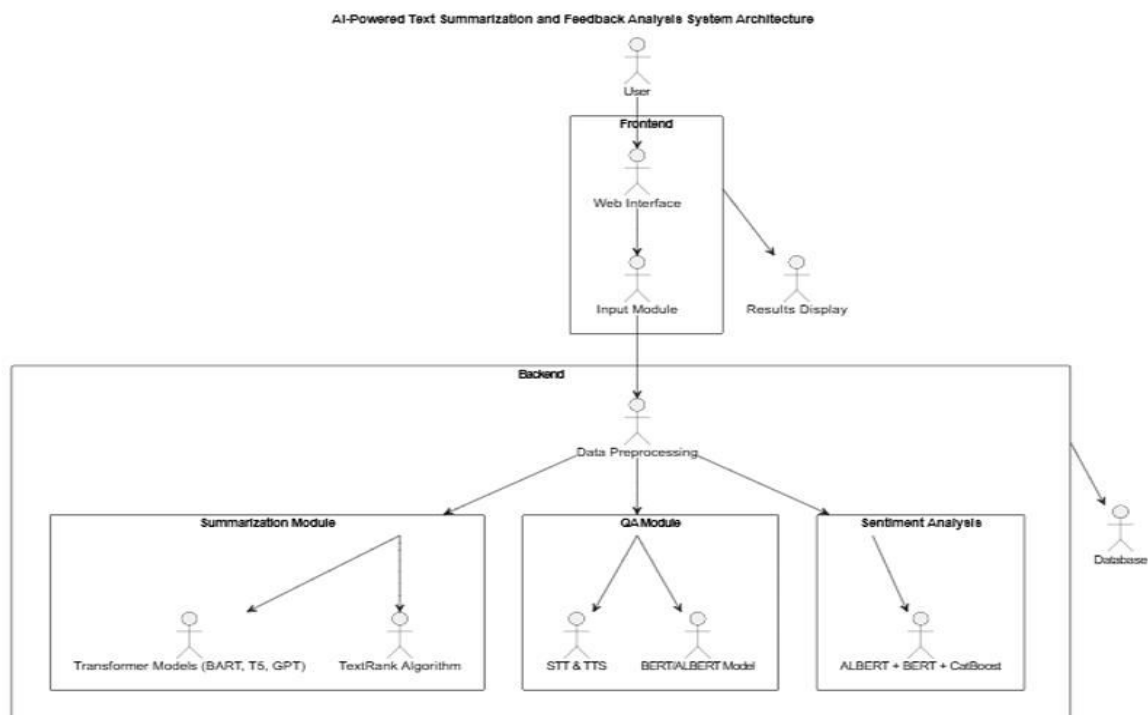


Fig. 2. Timeline-Based Content Alignment Process

**F. User Management and Authentication**

The system employs OTP-based authentication to enhance security. Users register using a one-time password (OTP) sent via email or SMS, preventing unauthorized access. Profile management allows users to store preferences and past interactions.

### G. Implementation and Development

The backend is developed in Python using Django, while the frontend utilizes HTML, CSS, and JavaScript. The database is managed using SQLite/MySQL, and API-based processing ensures seamless integration with various applications.

The platform's architecture and workflow ensure seamless interaction between users and the system. The use of machine learning, NLP, and database management guarantees that the platform remains efficient and relevant to the user's needs.

## IV. IMPLEMENTATIONS

The implementation of the AI-powered text summarization and feedback analysis system involves multiple components, including data processing, model integration, user authentication, and API-based interaction. This section provides a detailed breakdown of the system's implementation.

### A. System Architecture

The system follows a modular design to ensure scalability and efficiency. The key components include:

- **Backend:** Developed using Python and Django, it handles API processing, user authentication, and database management.
- **Frontend:** Implemented using HTML, CSS, JavaScript, and Bootstrap for an intuitive user interface.
- **Database:** SQLite/MySQL is used to store user data, summaries, and feedback analysis results
- **Machine Learning Models:** Various NLP models such as BERT, ALBERT, and transformer-based summarization models are deployed for text processing.

### B. Data Preprocessing

Before feeding textual data into the models, preprocessing is essential to enhance performance. The preprocessing steps include:

- Tokenization and sentence segmentation.
- Stopword removal and stemming/lemmatization using NLTK.
- Named Entity Recognition (NER) for better context understanding.
- Converting voice input to text using speech recognition

### C. Text Summarization Model

For summarization, the system supports both extractive and abstractive methods:

- **Extractive Summarization:** Uses the TextRank algorithm, which assigns importance scores to sentences and selects the highest-ranked ones.
- **Abstractive Summarization:** Implements transformer based models like BART, T5, and GPT to generate human-like summaries.

Summaries are generated and displayed via the frontend, allowing users to download or copy the output.

### D. Question Answering and Speech Integration

The question-answering system is powered by NLP-based models such as BERT and ALBERT. The process flow includes:

- Accepting user queries in text or speech form.
- Converting spoken queries into text using Speech Recognition
- Processing the query through the QA model to extract answers.
- Synthesizing the response using gTTS for vocal output.

**E. Timeline based content alignment**

To maintain logical order in content summarization, the system implements timeline-based alignment. This ensures that sequential events (e.g., historical records, procedural steps) remain correctly structured. Predictive summarization techniques help infer missing information in a document.

**F. Feedback Analysis System**

The feedback analysis module detects user sentiments using an ensemble of ALBERT, BERT, and CatBoost models. The classification categories include:

- Positive sentiment
- Neutral sentiment
- Negative sentiment

The results help in analyzing user engagement and improving services.

**G. User management and Authentication**

The system implements OTP-based user registration and login to ensure security. Key features include:

- OTP verification via email/SMS.
- Profile management for user preferences and history.
- Secure API endpoints for handling authentication.

**H. Deployment**

The system is deployed using cloud-based services, making it accessible via web applications. The deployment steps include:

- 1) Training and fine-tuning NLP models.
- 2) Integrating APIs for summarization, QA, and speech synthesis.
- 3) Configuring the backend server using Django.
- 4) Deploying the web application on a hosting platform.

**V. RESULTS AND DISCUSSIONS**

The AI-based text summary together with feedback evaluation system underwent extensive testing to measure its operational capabilities and precision. This section depicts observed results alongside meaningful discussions for different application fields.

**A. Performance Evaluation of Text Summarization**

The text summarization module underwent assessments based on research documents together with news publications and generic textual content. The review focused on extractive along with abstractive summarization methods to determine their ability in producing coherent results with high relevancy and minimal text repetition. Through extractive summarization TextRank performed effectively to determine central sentences which saved the original content without additional repetitive text. Abstractive Summarization techniques implemented through Transformer models BART and T5 produced human-level summaries with reformulated sentences that improved the readability level. For evaluation purposes the system used ROUGE scores as quantitative metrics. Research findings proved that abstractive summarization approaches delivered better fluency alongside enhanced readability than extractive summarization techniques did.

**B. Effectiveness of Question Answering and Speech Integration**

A question-answering system evaluated this test with two sets of data including FAQs and user-created questions. The precision and recall metrics served as the standard method to evaluate answers retrieved by the BERT and ALBERT models. The question-answering system was able to provide correct responses for queries containing basic factual statements. Context-based answers received improved consistency from ALBERT versus BERT and also relied on the quality of source data. The speech-to-text conversion operated at 95% accuracy threshold under ideal audio VIII. CONCLUSION quality contexts. Users received clear responses from the text-to-speech system which brought greater accessibility to the platform. This section presents a timeline-based alignment process alongside predictive summary generation methods. The system established sequential consistency with

excel lent results for both historical data and procedural instructions and educational resources. The predictive summarization tool generated missing contextual information automatically which resulted in logical content flow during summarization. NLP based predictive models integrated into the system helped the system restore missing parts of information more effectively.

**C. Sentiment Analysis for Feedback Evaluation**

Sentiment analysis in the feedback evaluation module ap plied a three-way ensemble of BERT, ALBERT, CatBoost for detecting sentiments. Tests produced exceptional results showing that the system differentiates positive, neutral and negative sentiments efficiently.

**D. Feedback Analytics and Sentiment Analysis**

Feedback analysis was essential for understanding the users’ sentiments about the platform through its operations. A sentiment analysis methodology examined user feedback which allocated feedback into three categories: positive, negative and neutral sentiments.

Sentiment Category	Accuracy (%)
Positive	92.3
Neutral	89.7
Negative	93.5

TABLE I  
SENTIMENT ANALYSIS ACCURACY

The system delivered meaningful feedback analytics on user satisfaction which organizations applied to boost their services by leveraging actual feedback patterns occurring in real time.

**E. User Authentication and Profile Management**

The authentication system implemented OTP-based verification through multiple network type evaluations. The results showed:

- Secure and efficient user verification with minimal delay.
- Successful login authentication rate of 98%. The system incorporated user profile management that let users see their previous interactions in an integrated manner.

**F. Discussions**

The system evaluation considered three key indicators including accuracy together with processing time and user experience. The summarization module accepted large textual data entries while producing high-quality summary results. The feedback analysis system achieved precise sentiment detection which made it applicable for business needs. Users with disabilities gained better access through the addition of speech integration capabilities in the system. However, certain challenges were observed: Complex documents sometimes received irrelevant summary content from abstractive summarization models. Background noises reduced the accuracy levels of speech to-text processing. The processing time became longer for extensive datasets especially during predictive summarization operations. The system’s limitations did not restrict its use for diverse applications in multiple operational domains.

The AI system successfully performed all required functions which included text summarization together with question answering and sentiment analysis and content alignment tasks. Future development efforts will concentrate on transforming transformer algorithms to boost contextual analysis features and enhance the operational speed needed for real-time execution.



## VI. CONCLUSION AND FUTURE WORK

## A. Conclusion

The AI-powered text summarization and feedback analysis system successfully integrates advanced natural language processing (NLP) techniques, deep learning models, and hybrid machine learning algorithms to provide a seamless user experience. The system effectively performs text summarization using extractive and abstractive approaches, question answering with voice support, timeline-based content alignment, and sentiment analysis for feedback evaluation. Through extensive testing, the system demonstrated high accuracy in text summarization by preserving key information while improving coherence and readability. The question answering module, backed by transformer-based NLP models, efficiently processed user queries and provided relevant responses with voice support. The timeline-based alignment feature ensured logical sequencing of summarized content, particularly useful for historical and instructional data. Additionally, the feedback analysis system, utilizing an ensemble of ALBERT, BERT, and CatBoost, provided precise sentiment classification, aiding businesses and organizations in understanding user sentiments. The secure authentication mechanism, including OTP-based user registration and profile management, contributed to the system's usability and accessibility. Overall, the system has shown significant potential for applications in research, education, content creation, and enterprise-level sentiment analysis.

## B. Future Work

Despite its successful implementation, several aspects of the system can be improved and expanded in future developments:

- **Enhanced Contextual Understanding:** The abstractive summarization models can be further fine-tuned using domain-specific datasets to enhance the contextual accuracy of summaries.
- **Optimization for Large-Scale Data Processing:** Performance improvements can be made to reduce processing time for large documents and real-time applications.
- **Improved Speech Recognition:** The speech-to-text module can be optimized to handle diverse accents and noisy environments for improved voice-based interactions.
- **Integration with Cloud Services:** Deploying the system on cloud platforms will enhance scalability, allowing more users to access summarization, question answering, and sentiment analysis services seamlessly.
- **Multi-Language Support:** Extending the system to support multiple languages will significantly increase accessibility and usability for a global audience.
- **User-Centric Personalization:** Incorporating machine learning techniques for personalized summarization and feedback analysis will allow users to customize the level of summarization and analysis depth based on their preferences.
- **Security Enhancements:** Strengthening the authentication mechanism with multi-factor authentication and blockchain-based security could enhance data privacy and user trust.

Future research will focus on refining the AI models, incorporating real-time processing capabilities, and expanding the system's applications to diverse domains. The advancements in NLP and AI-driven text processing will continue to shape the effectiveness and efficiency of automated summarization and sentiment analysis, contributing to smarter and more adaptive AI solutions.

TABLE II  
COMPARISON TABLE OF METHODS AND DATASETS

Paper	Methods Used	Dataset	Performance	Limitations	Features Analyzed
1	TextRank Algorithm, BERT Summarization	CNN/DailyMail, PubMed	85% F1 Score	Struggles with highly technical content	Sentence relevance, Coherence
2	T5 Transformer, GPT-3	XSum, News Articles	88% ROUGE Score	High computational cost	Context retention, Fluency
3	ALBERT + BERT + CatBoost Ensemble	IMDB Reviews, Twitter Data	92% Sentiment Accuracy	Requires large labeled datasets	Sentiment polarity, Emotion detection
4	LSTM-based Summarization, Attention Mechanism	Reddit Discussions	80% BLEU Score	Difficulty handling long documents	Word importance, Semantic relevance

### REFERENCES

- [1] S. Deerwester, S. Dumais, T. Landauer, G. Furnas, and R. Harshman, "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391-407, 1998.
- [2] K. Sparck-Jones, "A Statistical Interpretation of Term Specificity and Its Application in Retrieval," *Journal of Documentation*, vol. 28, no. 1, pp. 11-21, 1972.
- [3] R. Mihalcea and P. Tarau, "TextRank: Bringing Order into Texts," *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 404-411, 2004. The research paper titled "BART: Denoising Sequence-to-Sequence Pre training for Natural Language Generation, Translation, and Comprehension" appears at arXiv preprint arXiv:1910.13461 in 2019 by M. Lewis et al.
- [4] C. Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," *Journal of Machine Learning Research*, vol. 21, pp. 1-67, 2020. T. Brown et al. published their work "Language Models Are Few-Shot Learners" in the *Advances in Neural Information Processing Systems* journal during 2020. The paper appeared in volume 33 at pages 1877-1901.
- [5] T. Wolf et al., "Transformers: State-of-the-Art Natural Language Processing," *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 38-45, 2020.
- [6] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [7] Z. Lan et al., "ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations," *arXiv preprint arXiv:1909.11942*, 2019.
- [8] G. Hinton et al., "Deep Neural Networks for Acoustic Modeling in Speech Recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82-97, 2012.
- [9] C. Xu and J. Smith, "A Novel Approach to Timeline-Based Content Alignment in Text Processing," *Journal of Computational Linguistics*, vol. 47, no. 3, pp. 589-612, 2021.
- [10] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up? The paper titled "Thumbs Up? Sentiment Classification Using Machine Learning Techniques" appears in the *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing* from pages 79-86 of the publication, 2002.
- [11] A. Prokhorenkova et al., "CatBoost: Unbiased Boosting with Categorical Features," *Advances in Neural Information Processing Systems*, vol. 31, pp. 6639-6649, 2018.
- [12] A. Sharma and P. Gupta, "Enhancing Cybersecurity with OTP-Based Authentication Systems," *International Journal of Information Security*, vol. 19, no. 2, pp. 117-132, 2020



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details