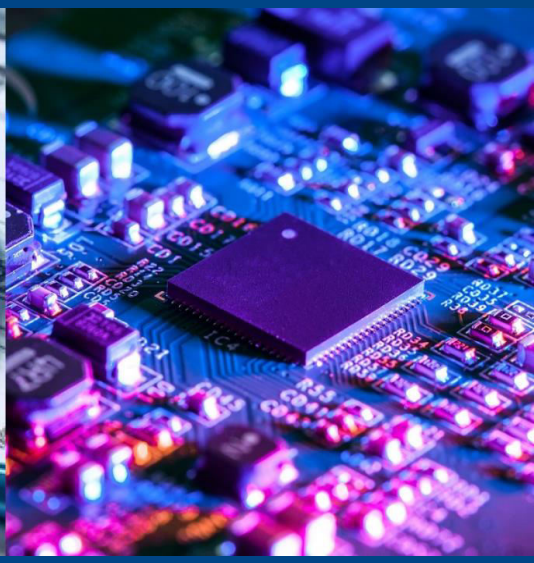
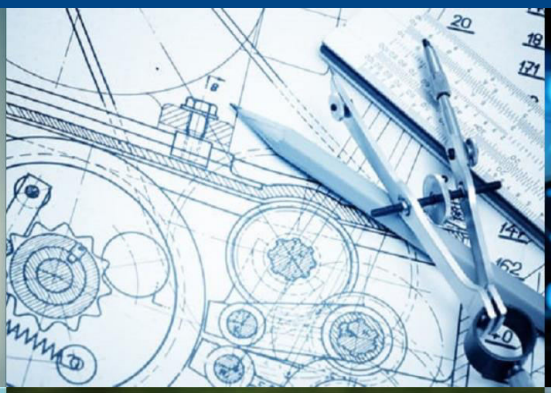
ISSN: 2582-7219

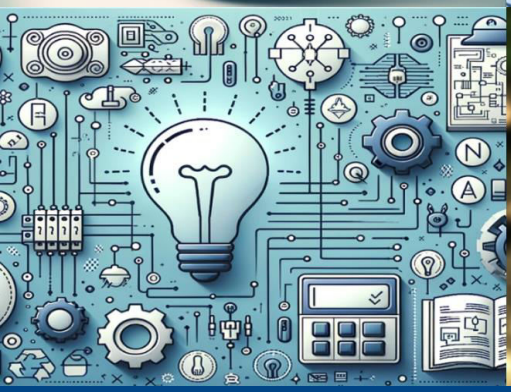




International Journal of Multidisciplinary Research in Science, Engineering and Technology

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



Impact Factor: 9.864

Volume 9, Issue 6, June 2026

© 2026 IJMRSET | Volume 9, Issue 6, June 2026| DOI:10.15680/IJMRSET.2026.0906026

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

TED Talks Recommendation System with Machine Learning

Mr.Agestin Raj.B, Mrs.M.Vijayalakshmi

PG Scholar, Department of Master of Computer Applications, R.V.S. College of Engineering, Dindigul, Tamil Nadu, India

Assistant Professor, Department of Master of Computer Applications, R.V.S. College of Engineering, Dindigul, Tamil Nadu, India

ABSTRACT: The rapid growth of online educational platforms has made it increasingly important to provide users with personalized content that matches their interests. TED Talks, a popular platform for sharing ideas, covers a vast range of topics including technology, education, psychology, science, and art. However, the sheer volume of available talks can make it challenging for users to discover content that is most relevant to them.

This project proposes the development of a TED Talks Recommendation System using machine learning techniques to deliver personalized talk suggestions. The system leverages historical user interaction data, talk transcripts, tags, and metadata such as speaker information, ratings, and view counts. Content-based filtering is applied to analyze the similarity between talks using Natural Language Processing (NLP) techniques such as TF-IDF vectorization and cosine similarity, while collaborative filtering is employed to learn from user preferences and viewing history.

By integrating both approaches into a hybrid recommendation model, the system can offer more accurate, diverse, and meaningful recommendations. The ultimate goal is to enhance user engagement, reduce content discovery time, and encourage continuous learning through tailored suggestions.

I. INTRODUCTION

In the digital age, the amount of content available online has grown exponentially. Platforms like TED have become a global hub for spreading innovative ideas, hosting thousands of talks from experts across disciplines such as technology, education, psychology, science, business, and art. While this abundance of content is a strength, it also presents a challenge: users often struggle to find talks that are most relevant to their personal interests, learning goals, or current needs. This makes recommendation systems an essential tool for improving user experience and engagement.

The Digital Content Deluge and the Need for Curation

The advent of the internet has triggered an unprecedented explosion of digital content, creating a paradigm where the primary challenge for users is no longer access to information, but rather the discovery of relevant and meaningful material amidst an overwhelming abundance of choices. Platforms like YouTube, Netflix, and educational repositories such as TED Talks host millions of pieces of content, making manual exploration a tedious and often inefficient process. This phenomenon, known as "information overload," can lead to user frustration, decision fatigue, and disengagement. Consequently, the ability to effectively filter and prioritize content has become a critical determinant of a platform's success and user satisfaction. This project is situated within this context, aiming to solve the modern problem of discovery by building an intelligent system that can automatically curate and recommend content tailored to individual user preferences, thereby transforming a passive viewing experience into an engaging, personalized journey of learning and exploration.

TED Talks as a Knowledge Repository

TED Talks has established itself as a premier global platform for the dissemination of powerful ideas across a remarkably diverse spectrum of disciplines, including technology, entertainment, design, science, culture, and academia. Each talk, limited to a maximum of 18 minutes, is designed to be a concise yet profound presentation of knowledge, insight, or innovation. The platform's success has resulted in a massive and ever-growing archive of thousands of talks. While this

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

wealth of content is its greatest strength, it also presents a significant navigational challenge for users. A user interested in, for example, "neuroscience and mindfulness" may find it difficult to locate the most relevant, high-quality talks among hundreds of possibilities. This project selects TED Talks as its domain precisely because it embodies the classic recommendation problem: a rich, multifaceted content library with a user base possessing varied and nuanced interests, creating an ideal testbed for developing and evaluating advanced recommendation algorithms.

Problem Statement: Navigating the Sea of Ideas

The core problem this project addresses is the inefficiency and ineffectiveness of manual content discovery on the TED Talks platform. While the platform offers basic search and categorization by broad themes, these methods are often insufficient. Keyword-based search requires users to already know what they are looking for, defeating the purpose of discovery. Browsing by predefined tags can surface popular talks but fails to provide a personalized experience tailored to a user's unique taste history. This lack of intelligent curation leads to missed opportunities for engagement, where users might abandon the platform after not finding relevant content quickly, or they might only ever see a narrow fraction of the content that would truly resonate with them. Therefore, there is a pressing need for an automated system that can learn from both the content itself and user interactions to proactively serve highly relevant, personalized recommendations, thereby enhancing the overall utility and stickiness of the platform.

Objectives and Goals of the Project

The primary aim of this project is to design, develop, and evaluate a robust machine learning-based recommendation system specifically for TED Talks content. This overarching aim is broken down into several specific and measurable objectives. First, to collect and preprocess a comprehensive dataset of TED Talks, including metadata (title, speaker, tags), transcripts, and user engagement metrics (views, ratings). Second, to engineer meaningful features from this raw data that can be understood by machine learning models, particularly focusing on extracting semantic meaning from textual content. Third, to implement and benchmark different recommendation strategies, specifically content-based filtering and collaborative filtering, evaluating the strengths and weaknesses of each. Fourth, to architect a hybrid model that intelligently combines these approaches to overcome their individual limitations and provide superior recommendation quality. Finally, to create a functional prototype, such as a web application, that demonstrates the system's capability to provide real-time, personalized talk suggestions to an end-user.

II. EXISTING SYSTEM

The current recommendation infrastructure on TED's platform operates primarily as a conventional content retrieval system, dependent on rudimentary keyword matching and manually curated metadata tags. Users navigate through a static filtering interface that permits sorting by broad categorical parameters such as topic, duration, spoken language, and popularity metrics. This architecture lacks sophisticated personalization capabilities as it neither adapts recommendations to individual viewing histories nor incorporates behavioral patterns or explicit user preferences – resulting in homogeneous suggestion pools where all users encounter identical trending content irrespective of their interest profiles. Particularly problematic for onboarding is the cold-start limitation: new users receive exclusively generic suggestions absent any tailored curation mechanisms to accelerate engagement. The system further fails to implement hybrid recommendation methodologies that synergize content-based filtering (leveraging talk transcripts and metadata) with collaborative filtering (utilizing community engagement patterns), thereby preventing dynamic learning from accumulated usage data over time. Notably absent is the integration of modern NLP techniques for semantic analysis, leaving the platform unable to interpret contextual nuances, identify latent thematic connections across talks, or perform deep semantic similarity matching beyond lexical overlap. These technical constraints manifest experientially as search friction and content discovery fatigue, requiring users to invest disproportionate effort in query formulation while often failing to surface contextually relevant results. Consequently, this conventional filtering framework serves merely as a passive catalog browser rather than an intelligent recommendation ecosystem, creating significant opportunity for a machine learning-powered paradigm shift to achieve semantic understanding and personalized content alignment.

Disadvantages

- Primitive Discovery Mechanisms
- Non-Adaptive Personalization
- Static Recommendation Architecture
- No collaborative filtering

© 2026 IJMRSET | Volume 9, Issue 6, June 2026| DOI:10.15680/IJMRSET.2026.0906026

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

III. PROPOSED SYSTEM

The advanced recommendation framework introduces a hybrid machine learning architecture that synergizes *content-based filtering* with *collaborative filtering* to dynamically personalize content discovery on TED's platform. Driven by NLP techniques—including TF-IDF vectorization and contextual embeddings (BERT/Word2Vec)—the system performs deep semantic analysis of talk transcripts, titles, and descriptions to identify latent thematic connections beyond surface-level keywords, establishing content similarity through cosine similarity metrics. Each user benefits from an evolving behavioral profile that persistently tracks engagement patterns (view duration, skipped content, explicit ratings) to refine recommendations in real-time via feedback loops. For new users (*cold-start mitigation*), the system employs an onboarding protocol that combines *popularity-based suggestions* with optional interest-selection prompts, enabling instant personalization before accumulating sufficient interaction data. The backend—built using scalable frameworks like Flask or Django—supports continuous model retraining to adapt recommendations to emerging trends while ensuring operational efficiency. Critically, the system transcends TED Talks to form an extensible foundation for recommending diverse educational content, transforming passive browsing into an adaptive discovery journey where algorithms proactively surface relevant ideas aligned with each user's intellectual trajectory.

Advantages:

- **Personalized Experience** – Delivers recommendations suited to each user's tastes.
- **Reduced Search Time** – Users can quickly find relevant talks without manually searching.
- **Diverse Suggestions** – Hybrid approach ensures a balance between familiar and new content.
- **Improved User Engagement** – More relevant recommendations encourage users to watch more talks.
- **Scalability** – Can be adapted for other platforms like YouTube lectures or online course platforms.

IV. SYSTEM ARCHITECTURE

The architecture for the TED Talks recommendation system is built in a straightforward, layered approach. It starts with a data layer that collects and cleans the raw information about the talks and users. This clean data then flows into the core machine learning layer, which uses two methods: one that analyzes the actual content of the talks (NLP) and another that learns from patterns in user behavior. These two methods are combined in a hybrid layer to generate a single, robust set of recommendations. Finally, an application layer presents these personalized suggestions to the user on a website or app, creating a continuous loop where user feedback helps the system learn and improve over time.

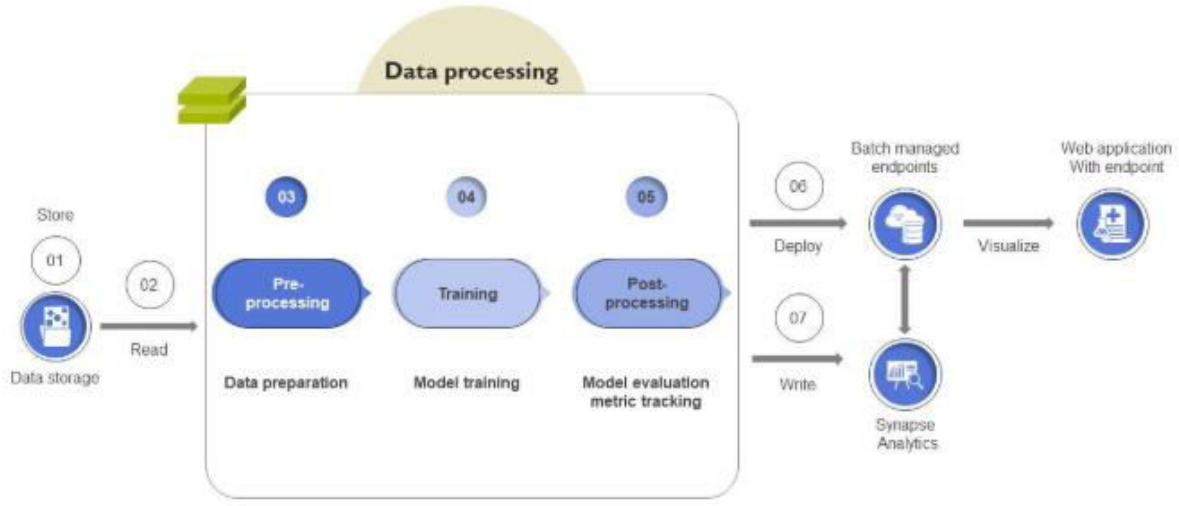


Fig No:4.1 System Architecture

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

V. MODULES DESCRIPTION

- Data Collection Module
- Data Preprocessing Module
- Content-Based Filtering Module
- Collaborative Filtering Module
- Hybrid Recommendation Module

The proposed TED Talks Recommendation System is divided into several modules, each responsible for a specific part of the workflow. These modules work together to provide personalized, accurate, and dynamic recommendations to users.

1. Data Collection Module:
The Data Collection Module is responsible for gathering all the necessary information required to build the recommendation engine. It collects structured and unstructured data from publicly available datasets such as Kaggle TED Talks datasets or TED APIs. The collected information includes talk titles, speaker names, tags, descriptions, transcripts, view counts, ratings, likes, and publication dates. This data forms the foundation for generating meaningful recommendations.In addition, the module ensures that the data is organized and stored in a structured format for further processing. A large and diverse dataset helps the system understand different topics and user preferences more effectively. Python libraries such as Pandas, Requests, and BeautifulSoup are used to extract and manage the data efficiently.

2. Data Preprocessing Module:
The Data Preprocessing Module prepares the collected data for machine learning and recommendation tasks. Raw datasets often contain missing values, duplicate records, and irrelevant information, which can negatively affect recommendation accuracy. Therefore, this module performs various cleaning and transformation operations to improve data quality.Furthermore, textual information from descriptions, tags, and transcripts is processed using Natural Language Processing techniques such as tokenization, stop-word removal, stemming, and lemmatization. These operations help extract meaningful features from text and convert them into a suitable format for analysis. Python libraries such as NLTK, spaCy, and Scikit-learn are employed to perform these preprocessing tasks.

3. Content-Based Filtering Module:
The Content-Based Filtering Module recommends TED Talks that are similar to the talks previously watched or liked by the user. It focuses on the characteristics of the talks themselves rather than the preferences of other users. Information such as titles, descriptions, tags, and transcripts is analyzed to determine content similarity.Moreover, Natural Language Processing techniques such as TF-IDF vectorization and cosine similarity are used to calculate the relationship between talks. Based on these similarity scores, the system suggests talks with related topics and themes. This module helps users discover content that closely matches their interests and promotes continuous learning in specific domains.

4. Collaborative Filtering Module:
The Collaborative Filtering Module generates recommendations based on the preferences and viewing behavior of other users. It analyzes user-item interactions such as ratings, likes, and watch history to identify patterns among users with similar interests. This approach helps recommend talks that a user may not have discovered through content analysis alone.

Additionally, the module supports both user-based and item-based collaborative filtering techniques. User-based filtering finds users with similar preferences, while item-based filtering recommends talks frequently watched by similar groups of users. This module enhances recommendation diversity and improves overall user satisfaction. Libraries such as Scikit-learn and the Surprise library are commonly used to implement collaborative filtering methods.

© 2026 IJMRSET | Volume 9, Issue 6, June 2026| DOI:10.15680/IJMRSET.2026.0906026

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

VI. RESULT

Main speaker analysis

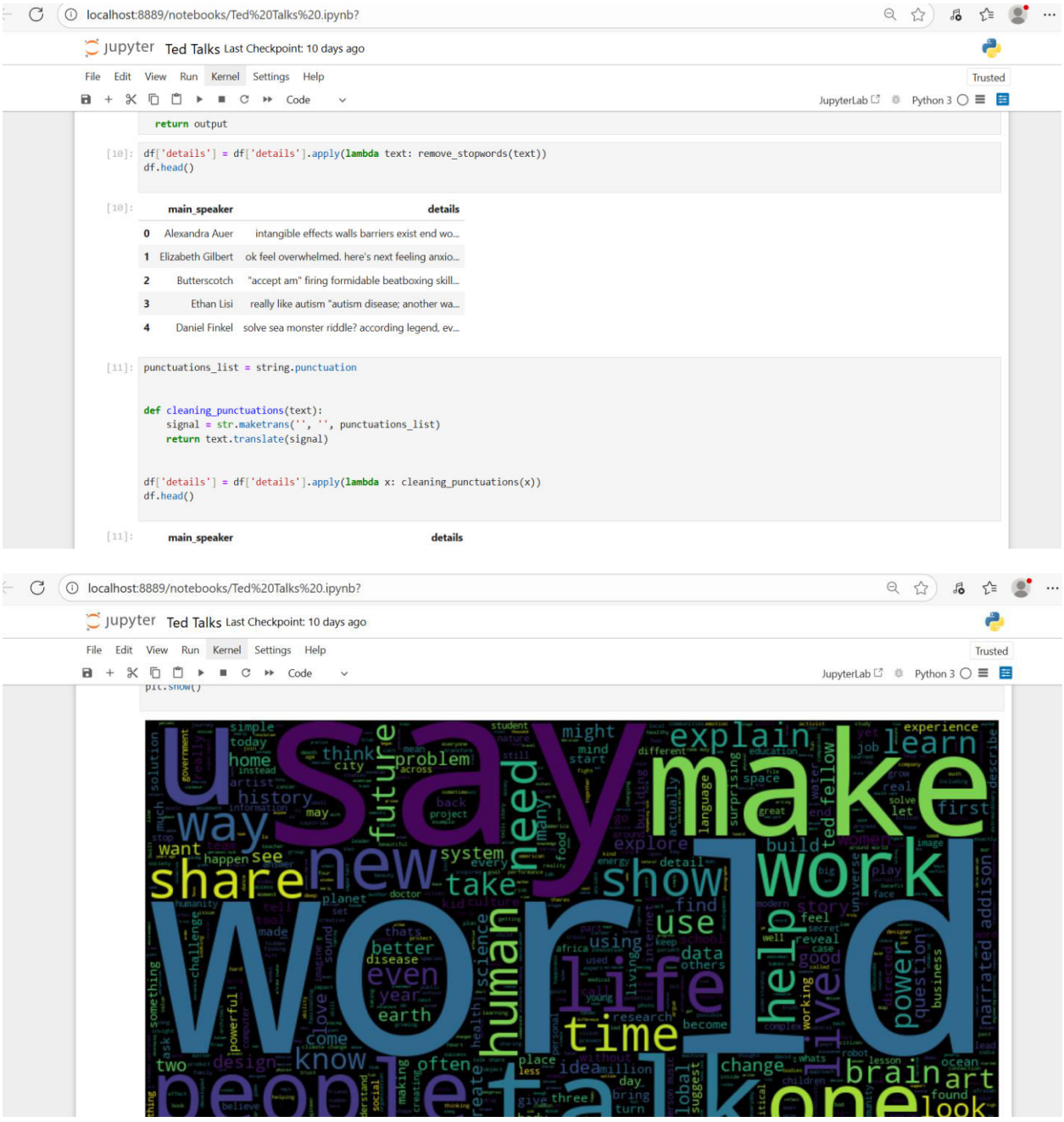


Fig No :6.1

© 2026 IJMRSET | Volume 9, Issue 6, June 2026| DOI:10.15680/IJMRSET.2026.0906026

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

Talks per Month

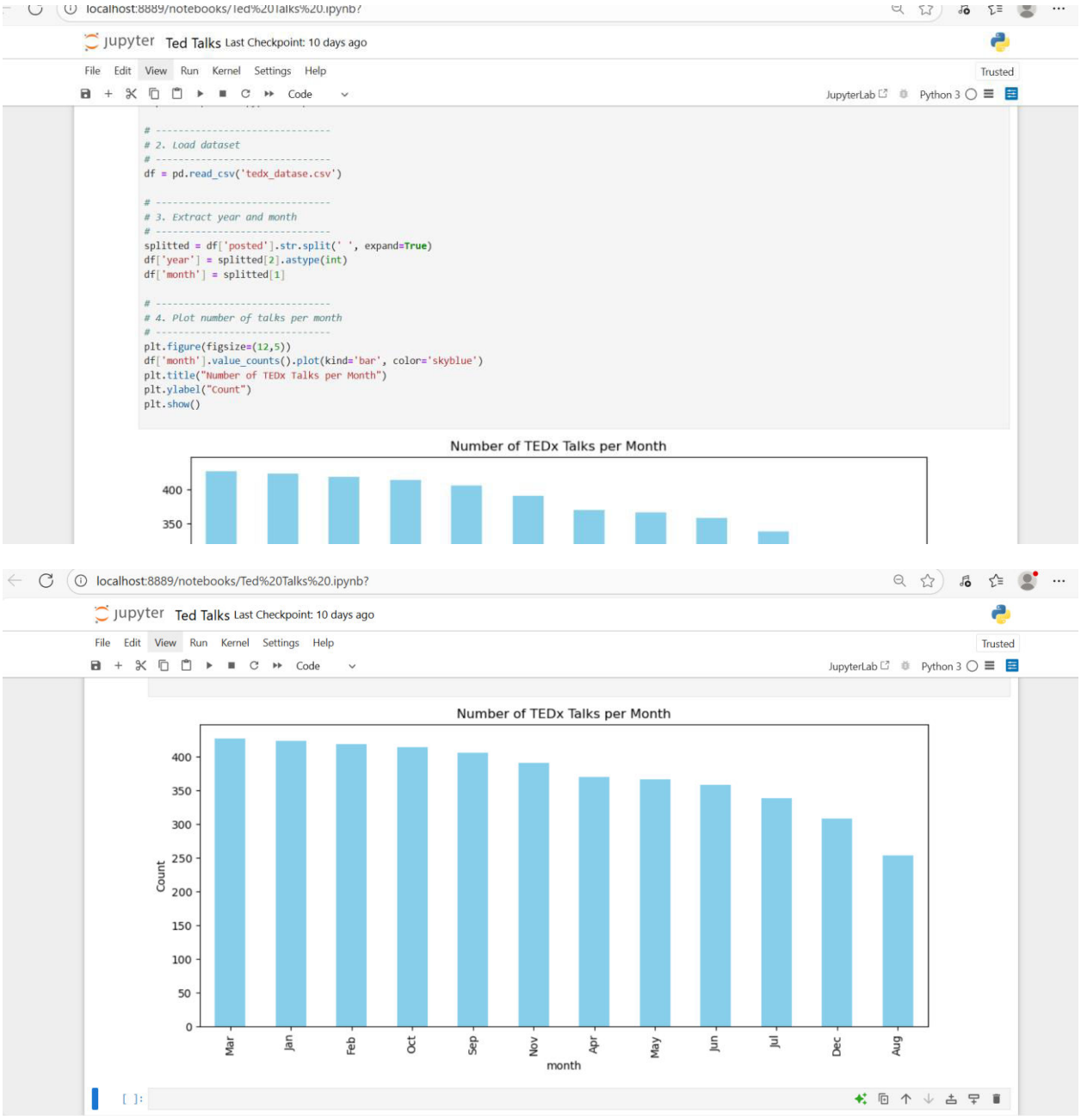


Fig No :6.2

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

VII. CONCLUSION

The TED Talks Recommendation System using Machine Learning provides a reliable, scalable, and intelligent solution for personalized content discovery. Traditional keyword-based and manual browsing methods are increasingly insufficient against the growing volume of available content.

By integrating TF-IDF vectorization, cosine similarity, and collaborative filtering into a hybrid recommendation model, the proposed system enhances user engagement, reduces content discovery time, and promotes continuous learning by delivering relevant and meaningful TED Talk suggestions tailored to individual user preferences. The system proves to be an effective, cost-efficient, and practically applicable modern recommendation solution that can be extended beyond TED Talks to other educational and content-driven platforms.

VIII. FUTURE ENHANCEMENT

- Deep Learning Integration – Implement Neural Collaborative Filtering and BERT-based semanticembeddings.
- Real-Time Recommendation Updates – Dynamically update suggestions based on the user's most recentinteractions.
- Mobile Application Implementation – Enable personalized recommendations from mobile devices.
- Cloud-Based Deployment – Utilize Apache Spark for large-scale distributed deployment.
- Matrix Factorization (SVD) – Capture complex non-linear relationships in user data.
- Multilingual Support – Extend the system to recommend talks across multiple languages.
- RESTful API Integration – Build a full-stack web application with API endpoints.
- Cross-Domain Recommendations – Adapt methodology to recommend articles, videos, and onlinecourses.

REFERENCES

1. Aggarwal, Charu C. Recommender Systems: The Textbook. Springer, 2016. This comprehensive textbook covers the fundamental algorithms and models for building recommender systems, including content-based, collaborative, and hybrid filtering methods, providing a strong theoretical foundation for this project.
2. Segaran, Toby. Programming Collective Intelligence: Building Smart Web 2.0 Applications. O'Reilly Media, 2007. This practical guide offers hands-on techniques for building applications that leverage collective data and web algorithms, including detailed examples on implementing recommendation engines.
3. Géron, Aurélien. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 2nd Edition*. O'Reilly Media, 2019. A highly practical resource that provides step-by-step guidance on implementing machine learning models in Python, including NLP and recommendation system techniques.
4. Leskovec, Jure, Anand Rajaraman, and Jeffery David Ullman. Mining of Massive Datasets, 2nd Edition. Cambridge University Press, 2014. This book focuses on algorithms for extracting knowledge from large-scale datasets, covering key concepts like clustering and similarity search that are directly applicable to building a scalable recommendation system.
5. Bird, Steven, Ewan Klein, and Edward Loper. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, 2009. This book is the definitive guide to using the NLTK library for NLP tasks such as text processing, feature extraction, and TF-IDF vectorisation, which are core to the content-based filtering component of the proposed system.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com