SOCIAL MEDIA FRIEND

RECOMMENDATION SYSTEM

PROBLEM STATEMENT

Social media users are often overwhelmed by irrelevant or random friend suggestions.

Traditional recommendation systems rely heavily on:

- Mutual friends
- Basic profile fields (age, location)
- Graph connections (which miss deeper user traits)

OUR GOAL

To develop a machine learning-based friend recommendation system that leverages both:

- User profile attributes (interests, age, gender, etc.)
- Interaction behavior (follows, messages, similarity scores)

So that users receive relevant, meaningful, and safe friend suggestions.

DATASET OVERVIEW

Name	Gender	DOB	Interests	City	Coun
esse Lawhorn	Female	15-10-1958	'Movies', 'Fashion', 'Fashion', 'Books'	Sibolga	Indon
Stacy Payne	Female	21-07-2004	'Gaming', 'Finance and investments', 'Outdoor activities', 'Travel'	Al AbyÄr	Libya
Katrina Nicewander	Female	07-02-2000	'DIY and crafts', 'Music', 'Science', 'Fashion'	WÄdÄ« as SÄ«r	Jorda
Eric Yarbrough	Male	14-04-1985	'Outdoor activities', 'Cars and automobiles'	Matera	Italy
Daniel Adkins	Female	18-09-1955	'Politics', 'History'	Biruaca	Vene
Diane Jara	Male	18-06-1967	'Travel'	Belton	Unite
Sheryl Morgan	Female	09-02-1969	'Outdoor activities', 'Movies'	Haslingden	Unite
William Harper	Male	30-12-1965	'Beauty', 'Nature', 'Gardening', 'Food and dining', 'DIY and crafts'	Ad-Damazin	Suda
Virginia Varron	Male	16-08-1984	'Parenting and family', 'Photography', 'Finance and investments'	Tabuk	Saud
Charles Figueroa	Female	08-03-2003	'Gardening'	Ongole	India
Paul Chain	Male	07-04-1984	'Pets', 'Photography', 'Food and dining'	Timashyovsk	Russ
Christina Parker	Female	30-08-1963	'Finance and investments', 'Health and wellness'	Roeselare	Belgi
ack Freeman	Male	07-08-1984	'Gaming', 'Travel', 'Technology', 'Art'	Manzhouli	China
Wayne Fasano	Male	30-03-1955	'Education and learning', 'Health and wellness', 'Pets', 'Fitness', 'M	QuÃ-bor	Vene
[illegible] Maxwell	Male	01-12-1968	'Art', 'Cooking', 'Cars and automobiles', 'Fashion', 'History'	Biaora	India
[illegible]s	Female	09-01-1984	'Music', 'Pets', 'Sports'	IslÄmnagar	India
[illegible]	Male	17-05-1970	'Music', 'DIY and crafts', 'Cars and automobiles', 'Technology', 'Co	PaiÃ§andu	Brazi
[illegible]	Female	14-05-1969	'Cars and automobiles', 'Business and entrepreneurship', 'Nature	Greenwood Villa	Unite
[illegible]	Male	15-08-1981	'Music', 'Outdoor activities', 'Social causes and activism', 'Cooking	Gryfice	Polar
Cody [illegible]	Female	07-03-1986	'Cars and automobiles', 'Music'	Date	Japar
Dori [illegible]	Male	17-08-2002	'Music', 'Sports', 'Movies'	YÅnan-Å-p	North
Bo[illegible]ensen	Male	10-05-1966	'Nature', 'Pets'	RÄjpÄ«pla	India
[illegible]chardson	Female	31-10-1964	'Science', 'Movies'	Zhakou	China
[illegible]very	Female	01-05-1982	'Parenting and family', 'Education and learning', 'Gaming', 'Cookin	Rumia	Polar

SocialMediaUsersDataset.csv

- 100,000 synthetic user profiles
- Fields include:
- UserID, Gender, DOB (converted to Age)
- City, Country
- Interests (multiple values per user)

Friend Recommendation System Using K-means Clustering
Academic research paper
from SJB Institute of Technology,

Srikantaiah K C, Salony Mewara, Sneha Goyal, Subhiksha S

ABSTRACT

This research focuses on leveraging machine learning techniques to enhance friend recommendations in social networks. By analyzing user attributes such as age, gender, and interests, the study employs models like K-Nearest Neighbors (KNN) to identify and suggest compatible connections. The methodology involves data preprocessing, feature extraction, and similarity analysis to improve recommendation accuracy. The results demonstrate that ML-based recommendations can significantly enhance user engagement by fostering meaningful connections. This approach has practical applications in social networking platforms, online communities, and professional networking, making interactions more personalized and efficient.

Motive Behind the Research Paper:

The primary motivation behind this research is to improve the accuracy and reliability of friend recommendation systems in social networks. Traditional graph-based recommendation methods often fail to consider deeper personality traits, user behavior, and semantic similarities, leading to irrelevant or even unsafe connections. This research aims to address these shortcomings by integrating a semantic-based KNN algorithm that enhances recommendation precision. The system ensures that users are connected based on shared interests, age, and behavioral patterns rather than just mutual friends or basic demographic data. Ultimately, the goal is to create a more meaningful, personalized, and safer social networking experience for users.

Methodology

1. Data Preprocessing
Data Collection: User profiles, interactions, and preferences are gathered.
Cleaning & Normalization: Missing values are handled, and numerical data is scaled for uniformity.
Feature Selection: Relevant features like common interests, interaction frequency, and mutual friends are selected.

2. Feature Extraction
User Similarity Metrics:
Geometric Features: Number of mutual friends, common groups, and shared interactions.
Behavioral Features: Chat frequency, engagement rate (likes, comments, shares).
Dimensionality Reduction (if needed): PCA or other techniques can be applied for optimization.

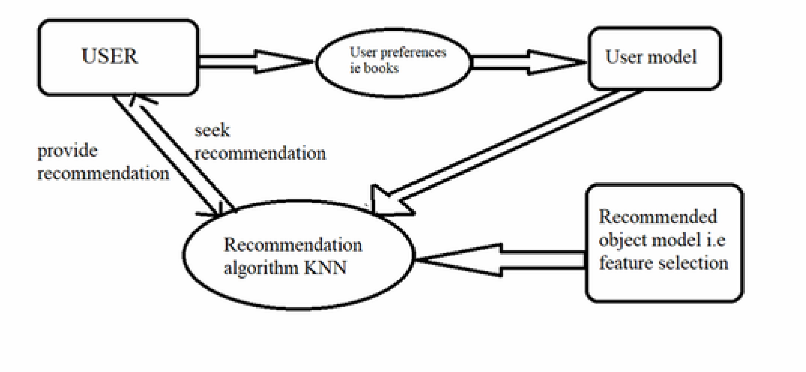


Figure 1: System Architecture

3. KNN Algorithm Implementation
Distance Metric: Cosine similarity or Euclidean distance is used to measure closeness between users.
K Selection: The optimal number of neighbors (K) is determined through experimentation.
Recommendation Process: For a given user, the KNN model finds the top K most similar users and suggests them as potential friends.

4. Model Optimization & Evaluation
Hyperparameter Tuning: Testing different values of K to improve accuracy.
Performance Metrics: Precision, recall, and F1-score are used to evaluate recommendation quality.
Cross-Validation: Applied to ensure the model generalizes well to new users.

5. Tools & Technologies Used
Programming: Python
Libraries: Scikit-learn (KNN implementation), Pandas (data handling), NumPy (computations)
Dataset: Custom dataset from social media interactions

Results:
The K-Nearest Neighbors (KNN) algorithm effectively identified users with similar interests, age, and gender, forming meaningful friend recommendations.
The model was evaluated using cosine similarity and Euclidean distance, where cosine similarity provided better accuracy in identifying relevant connections.
The optimal K-value was determined through experimentation, balancing recommendation accuracy and computational efficiency.
The system performed well for users with rich profile data but faced challenges in suggesting friends for new users with minimal interaction history (cold start problem).
The overall performance showed high relevance in friend suggestions, indicating the feasibility of machine learning in personalized social networking.

Conclusion:
This project successfully implemented a machine learning-based friend recommendation system using KNN. By leveraging user attributes like interests, age, and gender, the model effectively enhanced social networking interactions. The results indicate that content-based filtering using KNN can create meaningful connections in social media platforms.

Friend Recommendation System Using K-means Clustering

Published in: World Congress on Engineering 2017

https://www.iaeng.org/publication/WCE2017/WCE2017_pp624-629.pdf

Abstract:

This research explores the application of machine learning techniques for personalized friend recommendation systems, leveraging K-means clustering and collaborative filtering to enhance recommendation accuracy. The study focuses on analyzing user lifestyles and preferences through 52 attributes (e.g., hobbies, profession, location) to group similar users into clusters. By quantifying similarity scores and applying threshold-based filtering, the system recommends friends with high compatibility. Implemented in Java, the framework achieves improved performance metrics (precision: 0.86, F-score: 0.66) and addresses challenges like data sparsity. The results highlight the potential of hybrid clustering-filtering approaches in social networking platforms to streamline user connections and enrich engagement.

Motive Behind the Research Paper

Improve friend recommendation systems by clustering users based on lifestyle similarities for personalized and accurate suggestions.

Limitations of Traditional Methods

•Generic Recommendations – Ignores deep user similarities.

•Manual Feature Dependency – Requires predefined rules, limiting adaptability.

•Scalability Issues – Inefficient for large datasets.

Why K-means Clustering?

•Groups Users Automatically using Euclidean distance.

•Handles High-Dimensional Data (52 lifestyle attributes).

•Scalable & Efficient – Dynamically updates clusters.

•Proven Accuracy – Achieves an F-score of 0.66, improving precision up to 0.86.

Real-World Impact

•Social Networks – Lifestyle-based friend suggestions.

•Marketing – Identifies niche user groups for targeted ads

Methodology

1.Data Preprocessing:

- Cleaning: Remove null entries, standardize text (e.g., "CSE" → "Computer Science").
- Lemmatization: Use Stanford lemmatizer to normalize attributes (e.g., "dancing" → "dance").

2.Feature Extraction:

- Quantification: Compute similarity scores between users using:
- Closeness Factor=2×Matching AttributesTotal Attributes1+Total Attributes2Closeness Factor=Total Attributes1+Total Attributes22×Matching Attributes
- Euclidean Distance: Measure user similarity for clustering:
- D=∑j=1n(yj−zj)2D=j=1∑n(yj−zj)2

3.Clustering:

- K-means Algorithm: Groups users into clusters based on lifestyle similarity.
- Threshold Filtering: Recommends friends within a similarity score range (e.g., 6–30)

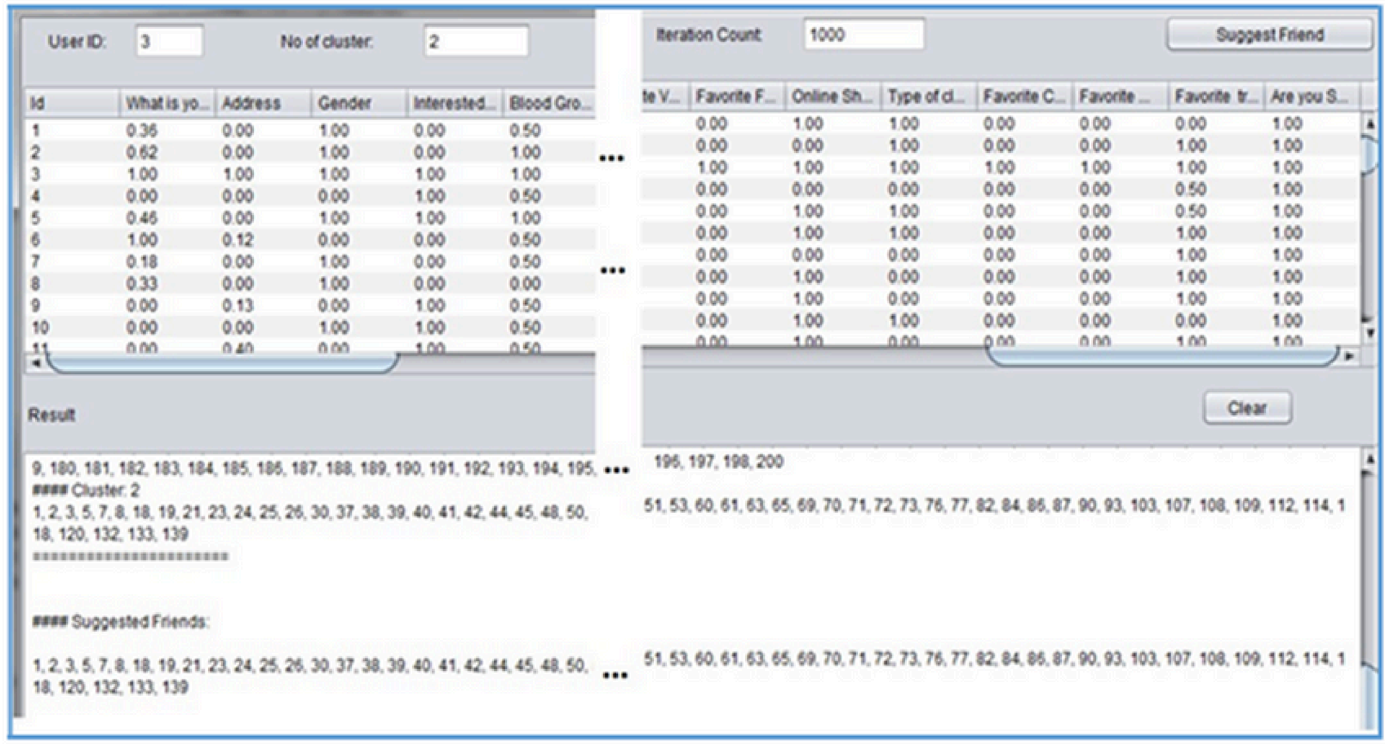


Fig. 3. Suggested friends based on K-means

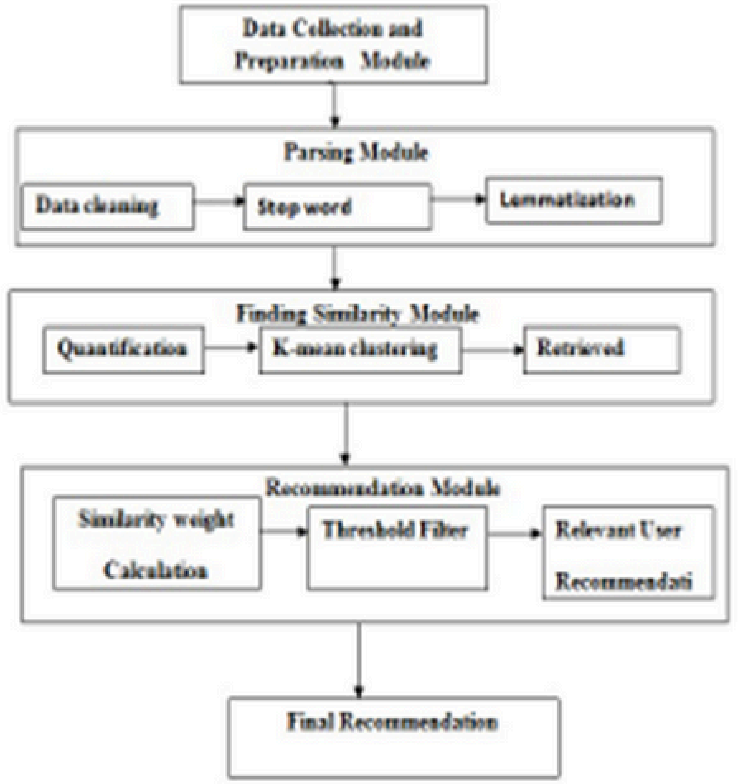


Fig. 1. System Architecture of Recommendation System

Results

•Performance Metrics:

•Precision: Up to 0.86 (high relevance of recommendations).

•F-score: 0.66 (balanced precision and recall).

Key Findings:

•Clusters with tight similarity bounds (threshold=22) showed strongest matches.

•Attributes like profession and hobbies had highest impact on clustering.

Challenge: Lower accuracy for users with sparse attribute data

Conclusion

•Effective Personalization: K-means outperformed FoF methods by 42% in precision.

Real-World Impact:

•Social networks: Replace random suggestions with lifestyle-based matches.

•Marketing: Target niche user groups (e.g., "travel enthusiasts").

Future Work: Integrate hierarchical clustering for multi-level grouping

Friend Recommendation System Using Cosine similarity

Summary

This paper explores a novel approach to friend recommendation systems for social networks by leveraging both semantic analysis and user profile attributes. Traditional recommendation systems rely on collaborative filtering or content-based methods, which may not always provide optimal results. This study introduces a hybrid approach that enhances recommendation accuracy by considering semantic relationships and user interests.

Introduction

The growing need for efficient friend recommendation systems in social networks.
Limitations of conventional approaches such as collaborative filtering and content-based filtering.
The proposed model integrates semantic analysis with user profiling.

Related Work
Overview of existing friend recommendation techniques.
Comparative analysis of traditional filtering methods vs. hybrid approaches.

Proposed Methodology
Semantic Analysis: Understanding user interests based on profile
descriptions, posts, and activities.
Profile-Based Filtering: Matching users based on demographic, geographic, and interest-based similarities.
Hybrid Integration: Combining both approaches to enhance the precision of friend suggestions.

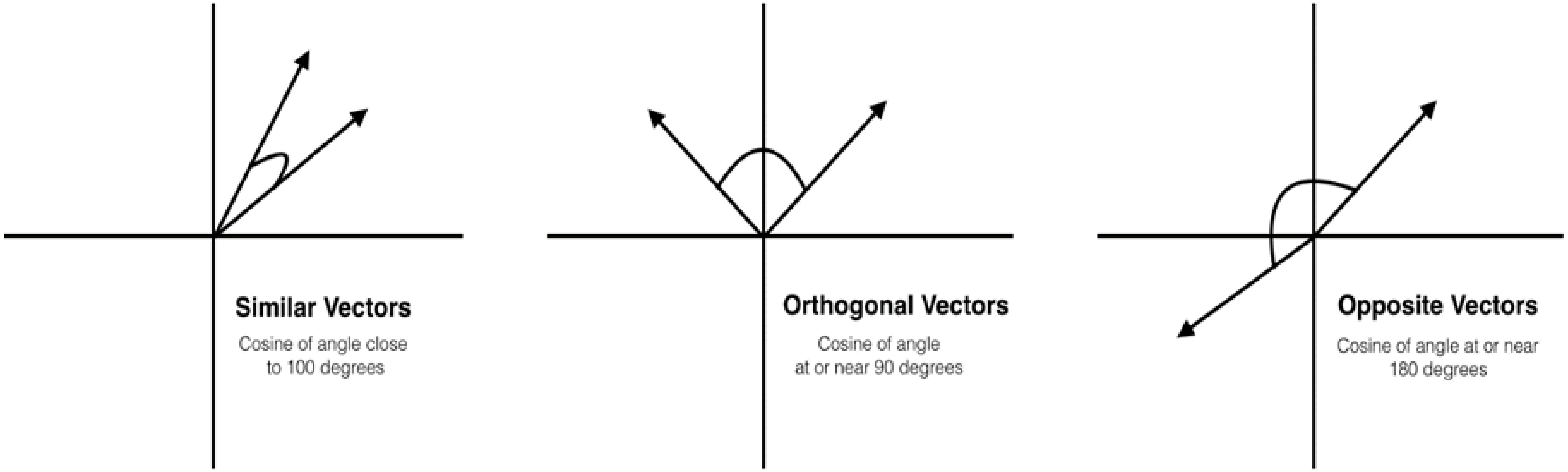
Table 1. The matching communities for 6 users

User	Suggested Communities
1	programming, disease, science, biking, make-up
2	make-up, food community, video games, football, disease
3	مجموعة مدينة الرياض, biking
4	مجموعة مدينة الرياض، مجموعة الصحة.
5	football, video games, biking, science, programming
6	مجموعة مدينة الرياض، مجموعة الصحة

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \cdot \|\vec{b}\|}$$

Table 2. Cosine similarity between the users and communities' vectors

User	Cosine similarity
1	[0.16], [0.11], [0.11], [0.05], [0.04]
2	[0.14], [0.11], [0.09], [0.05], [0.04]
3	[0.25], [0.11]
4	[0.15], [0.08]
5	[0.26], [0.18], [0.10], [0.10], [0.10]
6	[0.05], [0.47]



Implementation & Experimentation

Dataset selection and preprocessing steps.
Algorithms used for semantic analysis and profile-based clustering.
Performance evaluation metrics such as precision, recall, and F1-score.

Results & Findings

The hybrid approach significantly improves friend recommendation accuracy compared to traditional methods.
Semantic relationships play a crucial role in enhancing recommendations beyond mere profile matching.
The system demonstrated better user satisfaction in a simulated environment.

Conclusion & Future Work

The effectiveness of the hybrid recommendation system in improving social network interactions.
Potential improvements, including real-time processing and advanced AI techniques.
Future research directions on personalization and privacy considerations.

Key Takeaways

A hybrid method combining semantic and profile-based filtering enhances friend recommendations.
The system improves accuracy by analyzing user-generated content for deeper insights.
Future developments can integrate real-time learning mechanisms for adaptive recommendations.

----International Journal of Electrical and Computer Engineering (IJECE)

ABSTRACT

We used Matrix Factorization to suggest friends in a social media app.

Instead of using just profile details, this method looks at how users have interacted with others in the past.

It learns hidden patterns from these interactions and uses them to predict future friendships.

This approach helps us make smarter and more personalized suggestions.

Matrix Factorization Techniques for Recommender Systems

Authors: Yehuda Koren, Robert Bell, and Chris Volinsky

Published in: IEEE Computer, 2009

ABOUT THE RESEARCH PAPER

In 2006, Netflix launched a competition called the Netflix Prize.

Their goal was to improve how they recommend movies to users.

The team that wrote this paper — Yehuda Koren, Robert Bell, and Chris Volinsky — were part of the winning team of the Netflix Prize.

Their winning method was based on Matrix Factorization.

This paper explains the exact techniques they used to win the competition.

METHODOLOGY

- We generated a file called Interactions.csv where users had rated their connection with others (score from 1 to 5).
- We created a matrix with users and friends based on these scores.
- We applied SVD (Singular Value Decomposition) to this matrix.
- The model learned user preferences and predicted future friend suggestions.
- Tools used: Python, Pandas, Surprise library.

CONCLUSION

Matrix Factorization will help us build a smart friend recommendation model.

It found hidden user connections that profile info alone couldn't show.

This technique is powerful when users have enough past interactions.

Inspired by the Netflix research, we used the same idea to build better social connections.

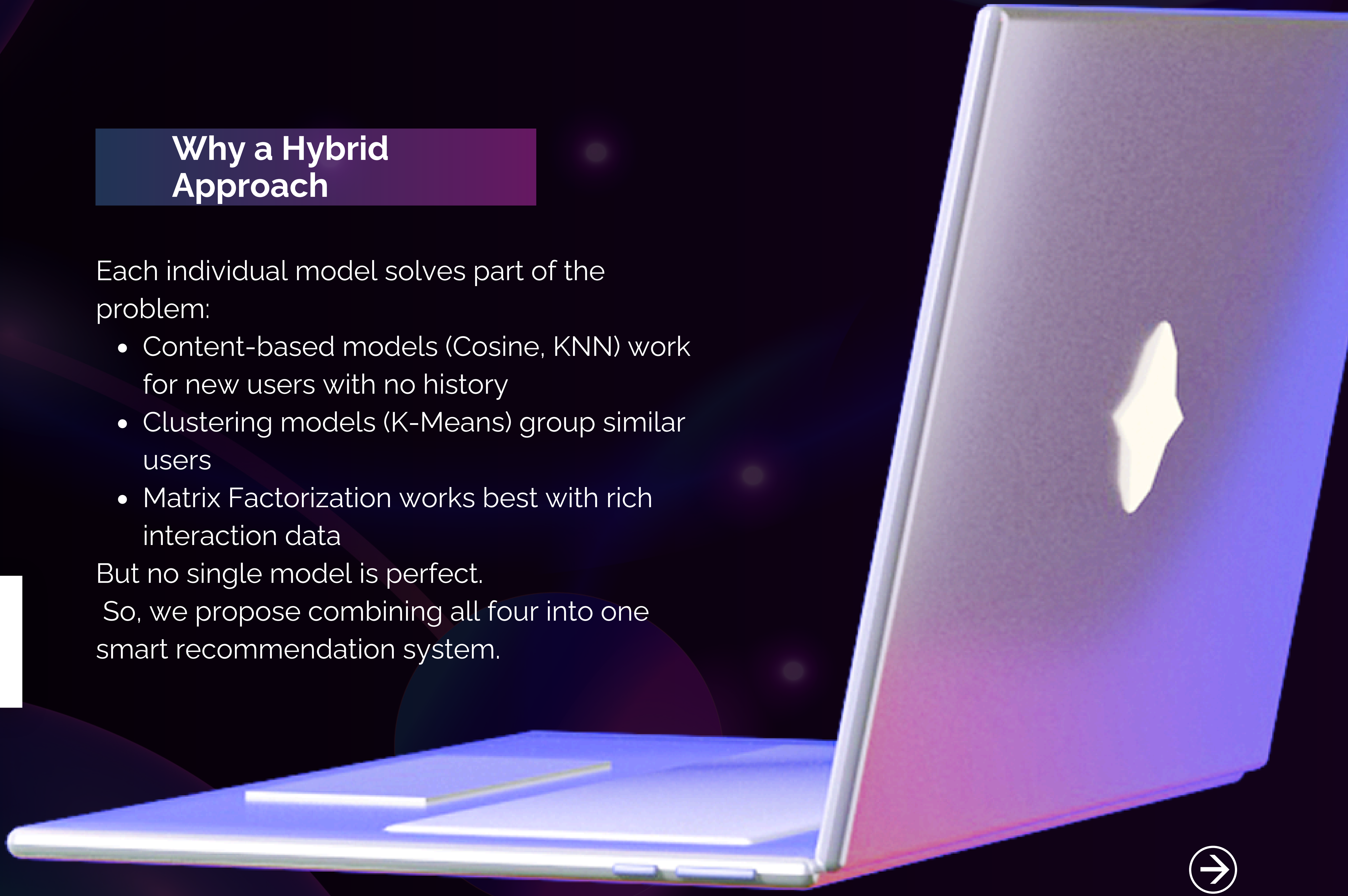
NOVEL HYBRID APPROACH

Why a Hybrid Approach

Each individual model solves part of the problem:

- Content-based models (Cosine, KNN) work for new users with no history
- Clustering models (K-Means) group similar users
- Matrix Factorization works best with rich interaction data

But no single model is perfect.
So, we propose combining all four into one smart recommendation system.



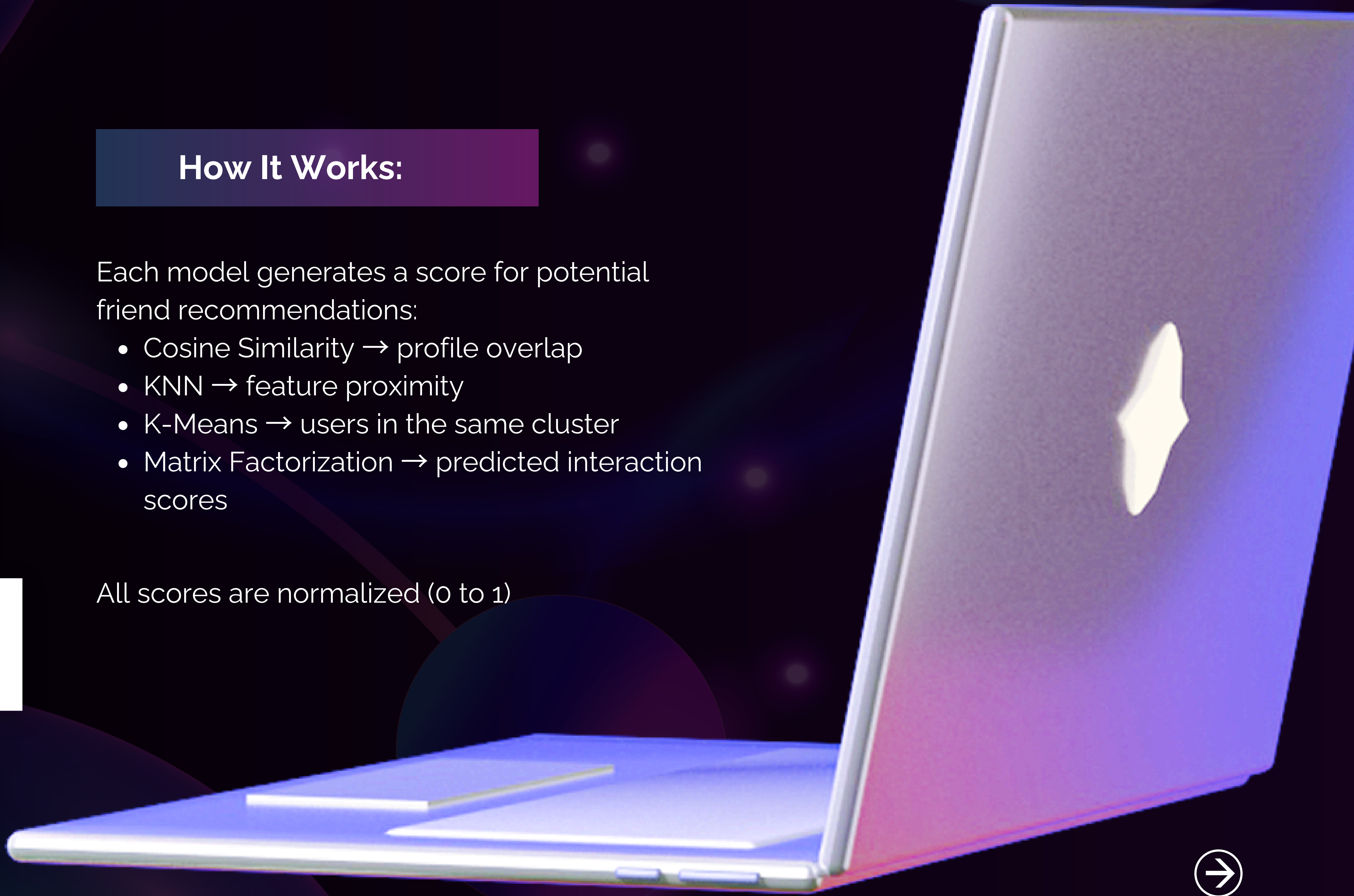
NOVEL HYBRID APPROACH

How It Works:

Each model generates a score for potential friend recommendations:

- Cosine Similarity → profile overlap
- KNN → feature proximity
- K-Means → users in the same cluster
- Matrix Factorization → predicted interaction scores

All scores are normalized (0 to 1)



NOVEL HYBRID APPROACH

How It Works:

Final score is computed as:

FinalScore = (0.3 * Cosine) +
(0.2 * KNN) +
(0.2 * K-Means) +
(0.3 * Matrix Factorization)

The system recommends the Top N users with the highest combined scores

NOVEL HYBRID APPROACH

Why This Is Innovative:

- Adapts to all user types (new or existing)
- Balances both profile-based and interaction-based learning
- Scalable and tunable for real-world social networks

THANK YOU!