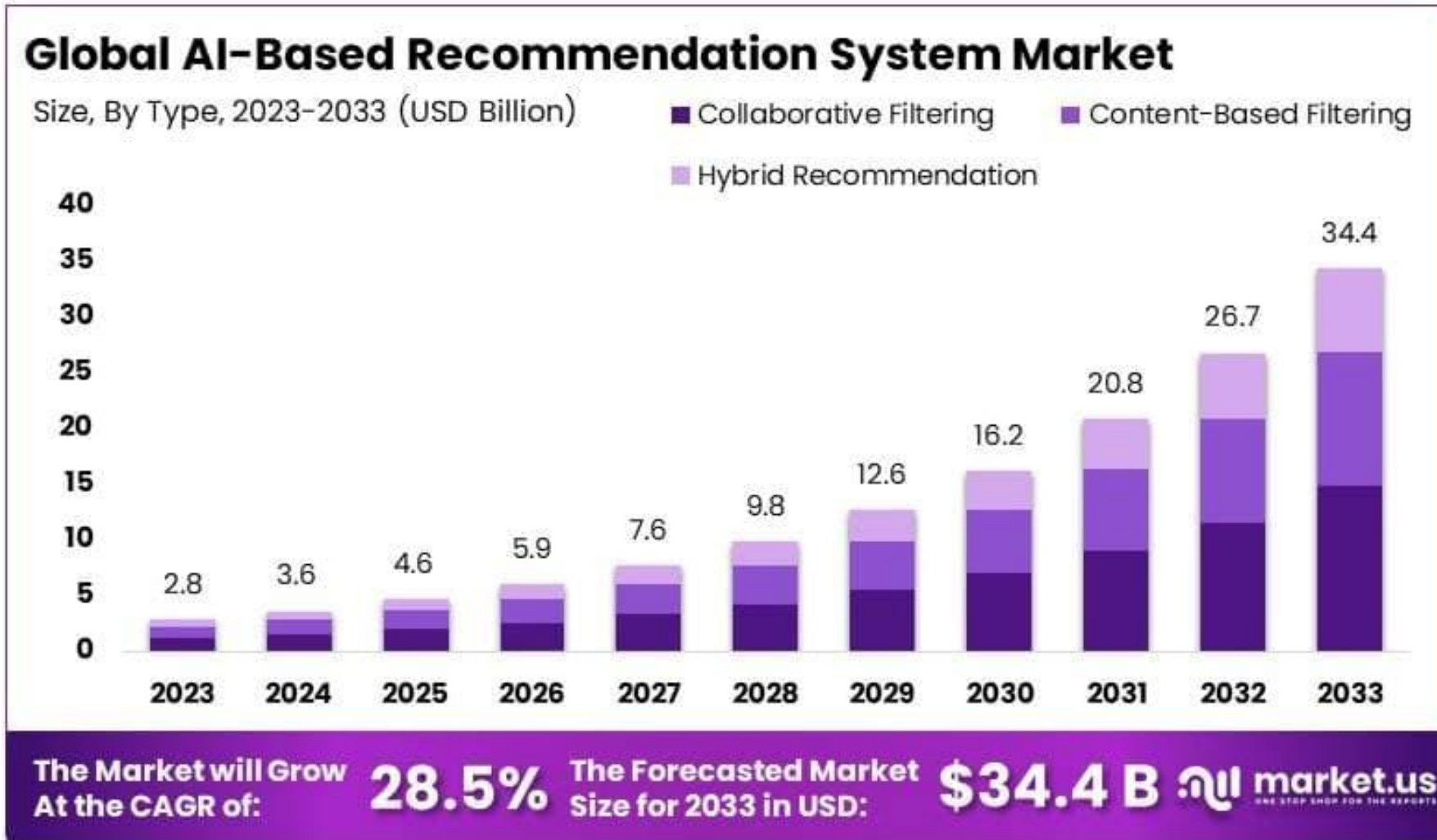




Recommendation System with Collaborative Filtering

Presenter: Tai Bui

RECOMMENDATION SYSTEM OVERVIEW



The global recommendation system market was valued at **USD 2.8 billion in 2023** and is projected to grow at a compound annual growth rate (CAGR) of **28.5% from 2023 to 2030**.

PROJECT OBJECTIVES

- o1. Develop a recommendation system using collaborative filtering
- o2. Minimizing model's root mean squared error





DATA OVERVIEW

- Provided by MovieLens, a movie recommendation service
- Over 100,000 ratings across 9742 movies
- Ratings are made on a 5-star scale, with half-star increments (0.5 stars - 5.0 stars)

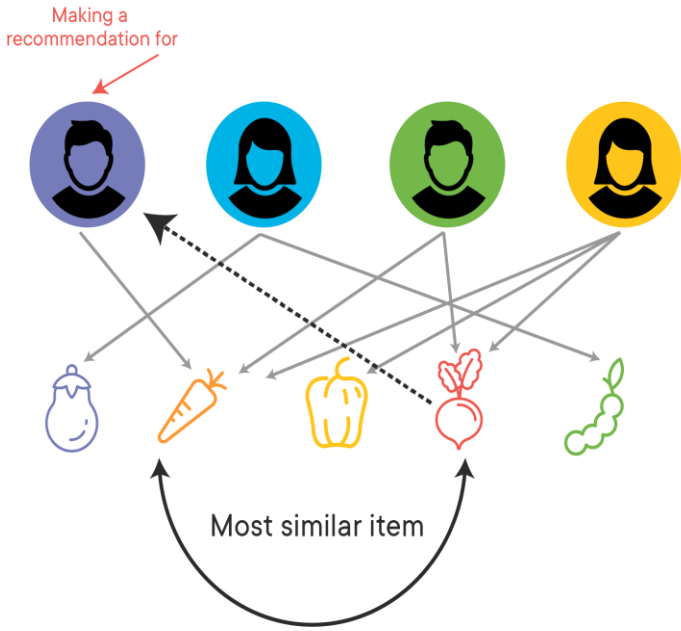
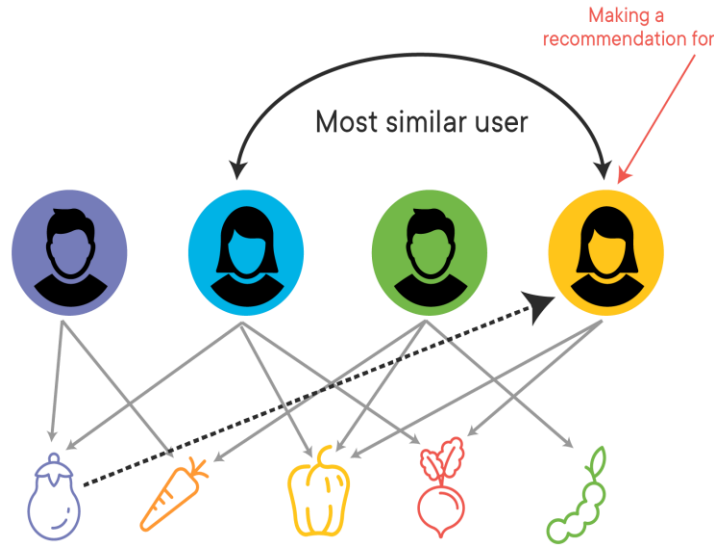


Diagram illustrating user-based collaborative filtering. Four users (purple, blue, green, yellow) are shown at the top. Below them are five items: a blue bottle, an orange carrot, a yellow pepper, a red radish, and a green pea. Arrows show connections from users to items. A curved arrow labeled "Most similar user" points from the blue user to the yellow user. A red arrow points to the yellow user with the text "Making a recommendation for".



- based on the similarity in preferences, tastes and choices of users
- There are 2 types: memory-based and model-based

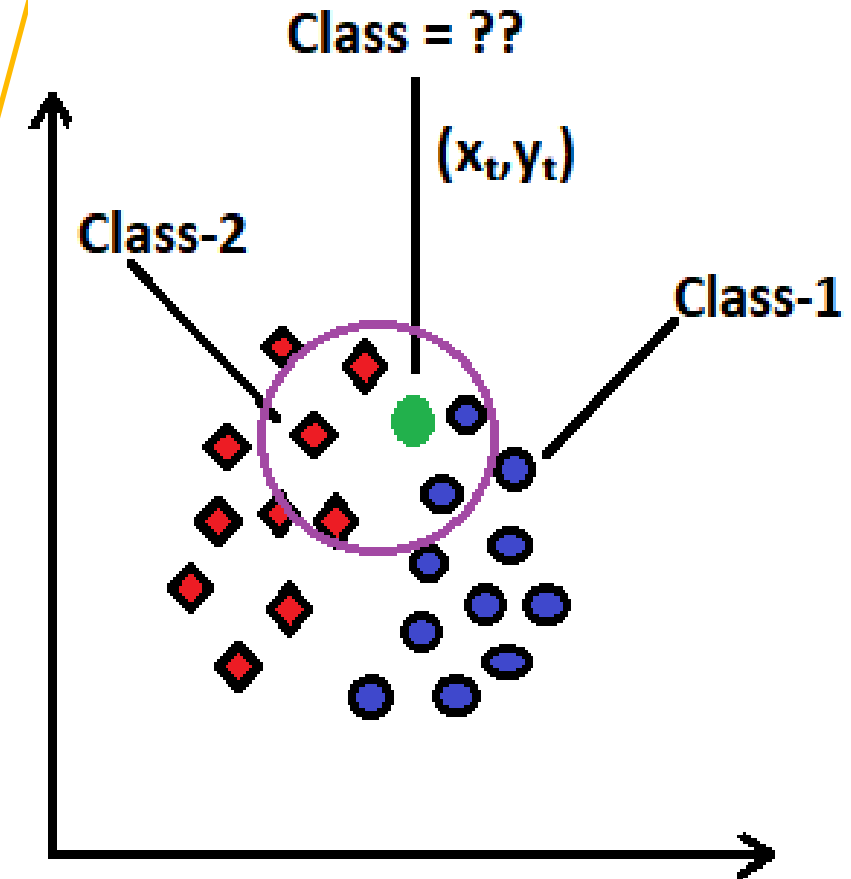
MEMORY-BASED VS. MODEL-BASED

Memory-based	Model-based
<ul style="list-style-type: none">▪ rely on the user-item interaction to make recommendations▪ Requires the user-item matrix to be available in memory▪ Limited scalability	<ul style="list-style-type: none">• learn patterns from the data and predict user preferences.• Uses the user-item matrix to train a model, then discards it after building the model.• Scales better for large datasets• Risk of overfitting



BASE MODEL - KNNBASIC MODEL

- simple, memory-based model that relies on measuring the similarity between data points to make predictions.
- 0.97 RMSE for both Pearson and Cosine similarity



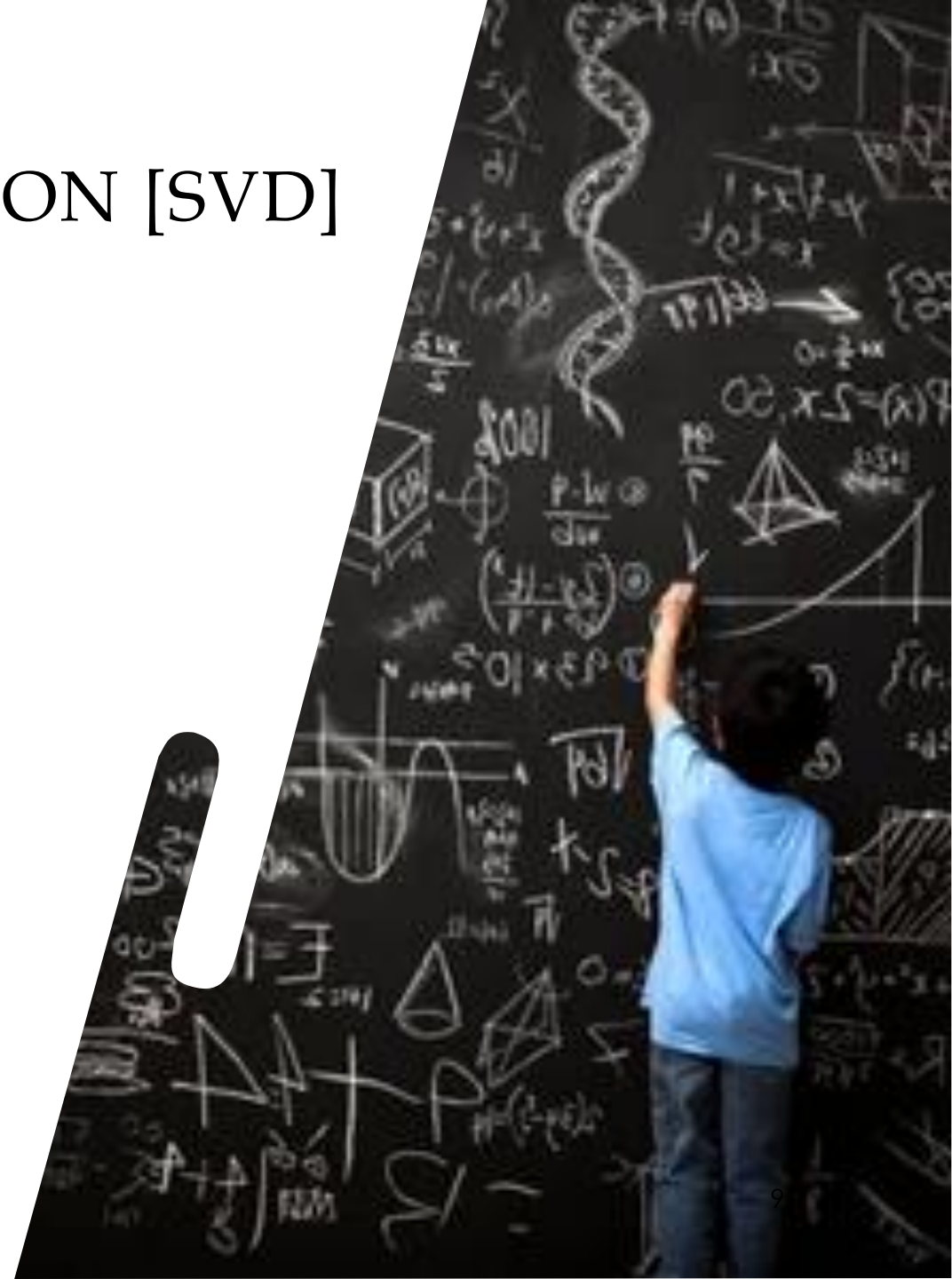
KNNBASELINE MODEL

- A memory-based KNN model adding in bias term to reduce overfitting
- RMSE of 0.8765 with Pearson similarity
- RMSE of 0.8782 with Cosine similarity



SINGULAR VALUE DECOMPOSITION [SVD]

- A model-based matrix factorization model
- A GridSearch is used to obtain best parameters
- The model with best params achieve 0.85 RMSE – the lowest out of the 3 models



CONCLUSION

- recommend Model-based Singular Value Decomposition (SVD) with user-based focus
- For future works, a hybrid model (content-based and collaborative filtering) will help solving the problem of new entries





Thank you

Please send all questions to:
taingocbui@gmail.com

Contact:
<https://www.linkedin.com/in/tai-bui-ngoc/>