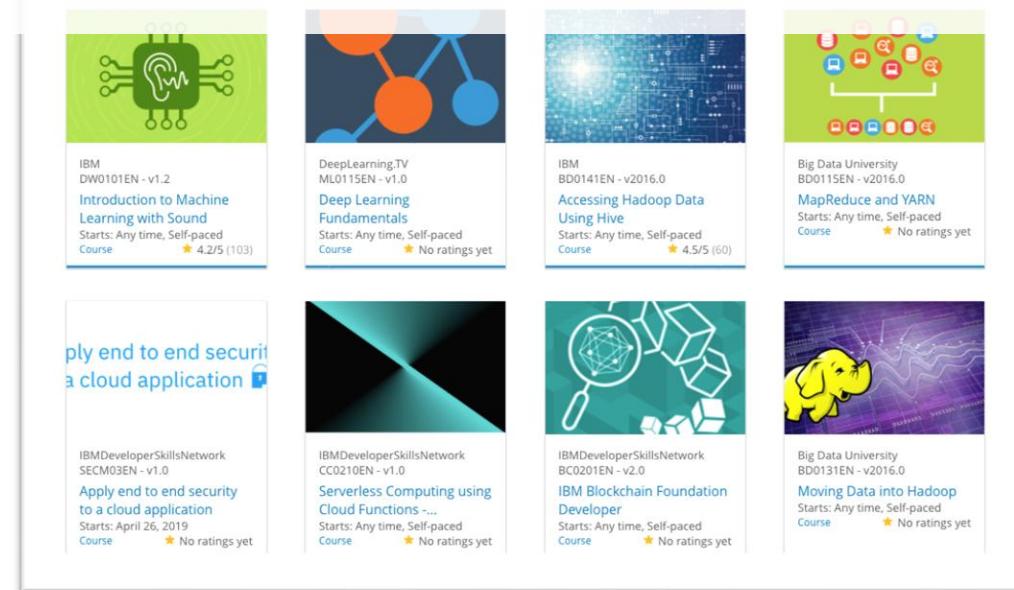


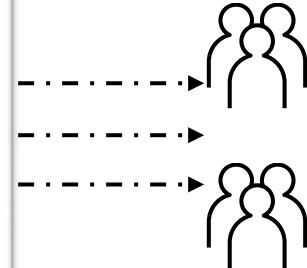
# Build a Personalized Online Course Recommender System with Machine Learning

Justin Matthew T  
**29 Nov 2025**



The grid displays nine course cards:

- IBM DW0101EN - v1.2**  
Introduction to Machine Learning with Sound  
Starts: Any time, Self-paced Course ★ 4.2/5 (103)
- DeepLearning.TV ML0115EN - v1.0**  
Deep Learning Fundamentals  
Starts: Any time, Self-paced Course ★ No ratings yet
- IBM BD0141EN - v2016.0**  
Accessing Hadoop Data Using Hive  
Starts: Any time, Self-paced Course ★ 4.5/5 (60)
- Big Data University BD011SEN - v2016.0**  
MapReduce and YARN  
Starts: Any time, Self-paced Course ★ No ratings yet
- IBMDeveloperSkillsNetwork SEC03EN - v1.0**  
Apply end to end security to a cloud application  
Starts: April 26, 2019 Course ★ No ratings yet
- IBMDeveloperSkillsNetwork CC0210EN - v1.0**  
Serverless Computing using Cloud Functions ...  
Starts: Any time, Self-paced Course ★ No ratings yet
- IBMDeveloperSkillsNetwork BC0201EN - v2.0**  
IBM Blockchain Foundation Developer  
Starts: Any time, Self-paced Course ★ No ratings yet
- Big Data University BD0131EN - v2016.0**  
Moving Data into Hadoop  
Starts: Any time, Self-paced Course ★ No ratings yet



# Outline

---

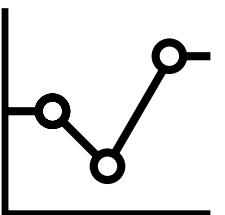
- Introduction and Background
- Exploratory Data Analysis
- Content-based Recommender System using Unsupervised Learning
- Collaborative-filtering based Recommender System using Supervised learning
- Conclusion
- Appendix

# Introduction

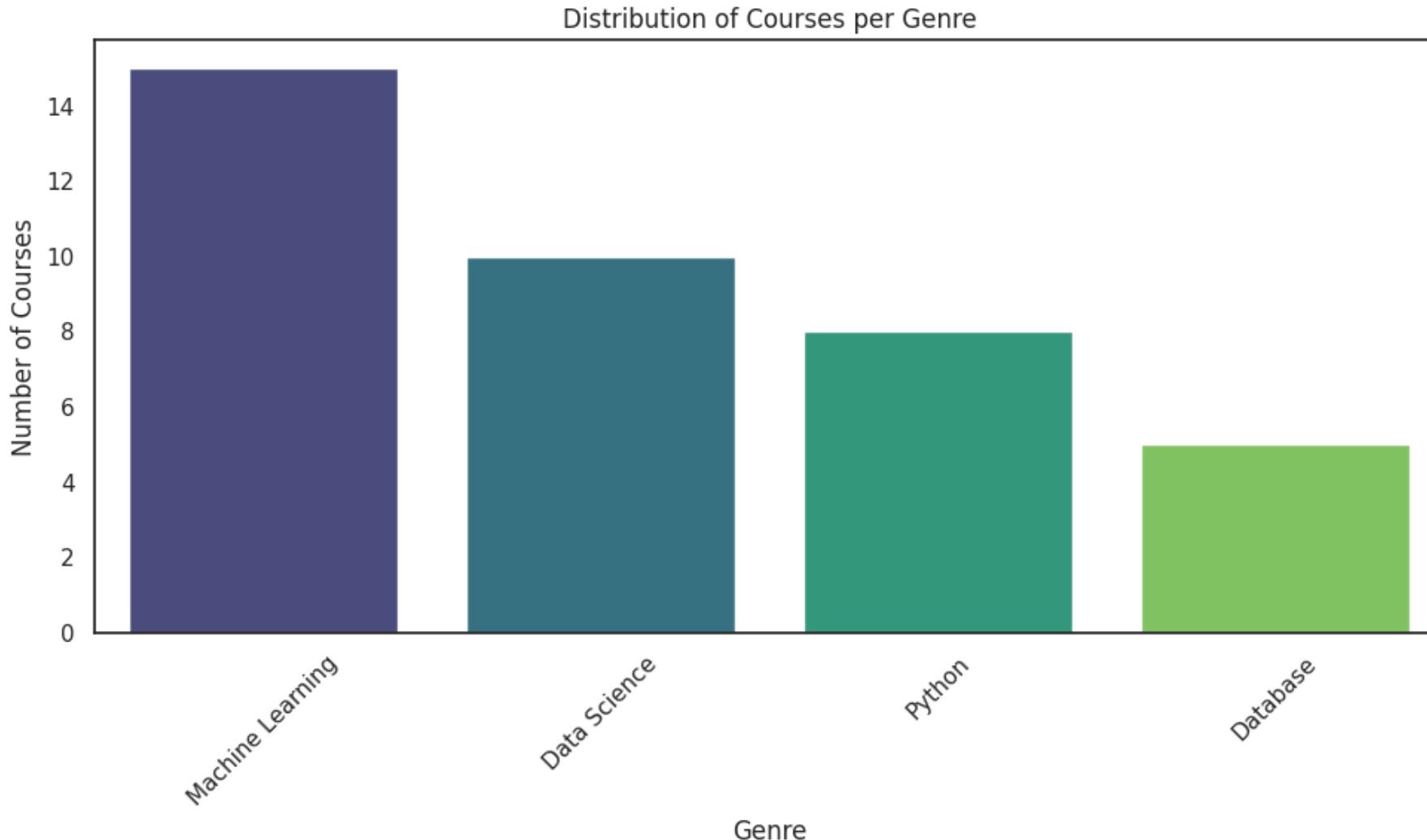
---

- Project background and context
  - Online education platforms host thousands of courses, leading to information overload for users.
- Problem states
  - Users struggle to find relevant courses. We need a system to personalize the learning experience based on user activities and preferences
- Hypotheses
  - **H1:** Users with similar genre interests will prefer similar courses.
  - **H2:** Neural Network embeddings can capture complex user-item interactions better than linear models.

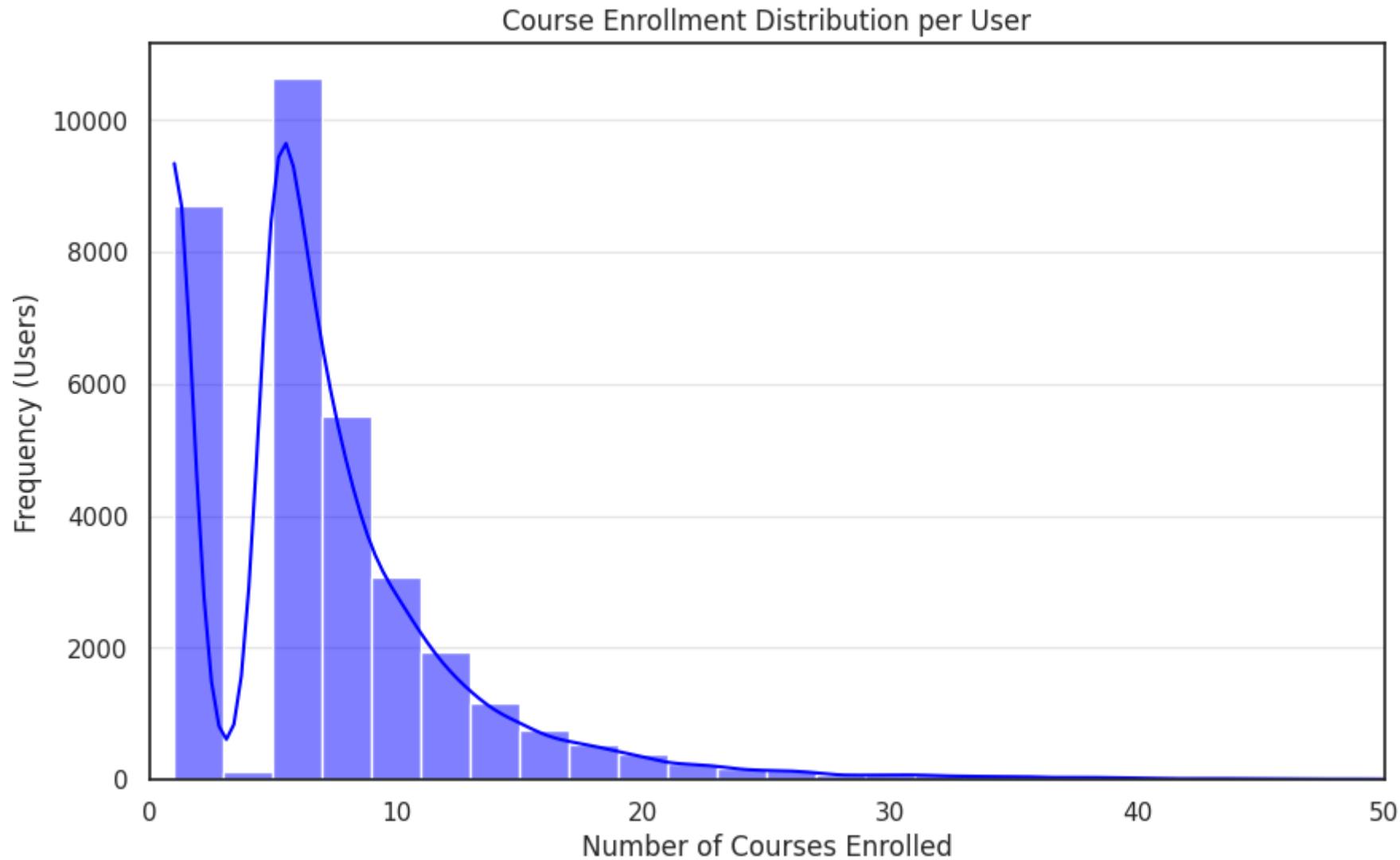
# Exploratory Data Analysis



# Course counts per genre



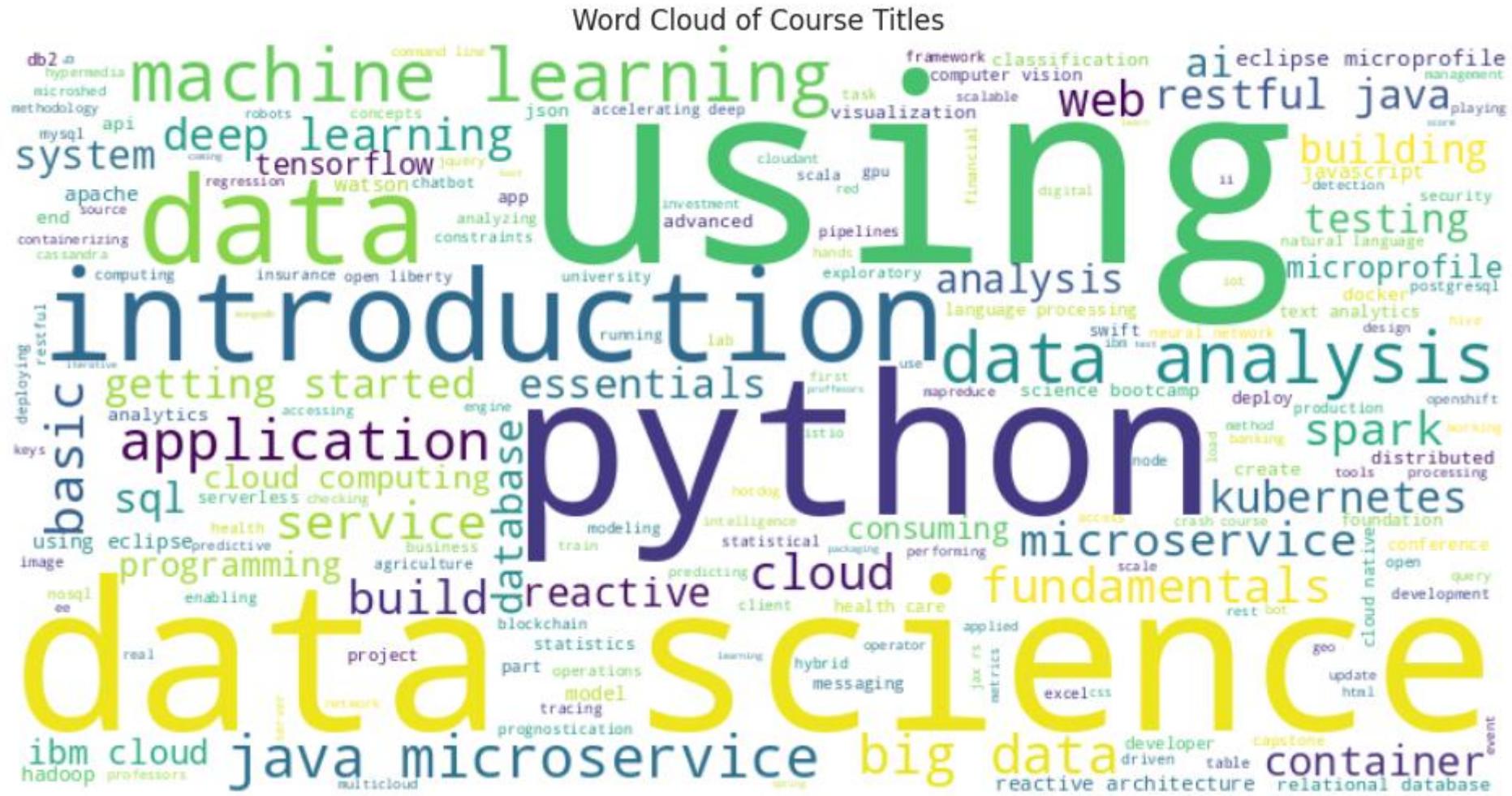
# Course enrollment distribution



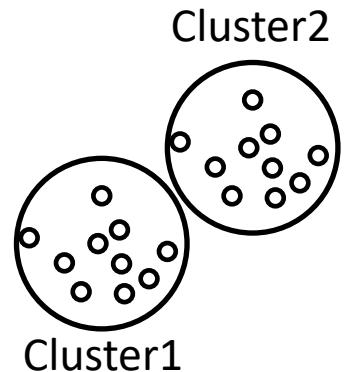
# 20 most popular courses

1. PY0101EN
2. DS0101EN
3. BD0101EN
4. BD0111EN
5. DA0101EN
6. DS0103EN
7. ML0101ENv3
8. BD0211EN
9. DS0105EN
10. BC0101EN
11. DV0101EN
12. ML0115EN
13. CB0103EN
14. RP0101EN
15. ST0101EN
16. CC0101EN
17. CO0101EN
18. DB0101EN
19. BD0115EN
20. DS0301EN

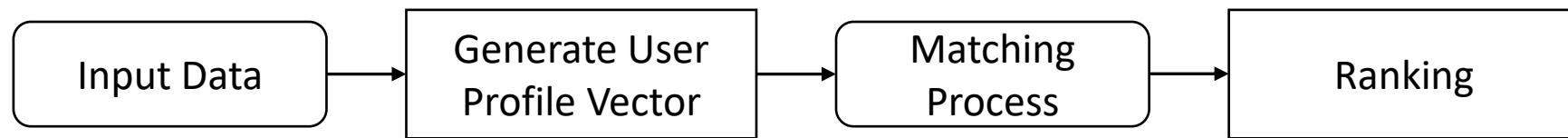
# Word cloud of course titles



# Content-based Recommender System using Unsupervised Learning



# Flowchart of content-based recommender system using user profile and course genres

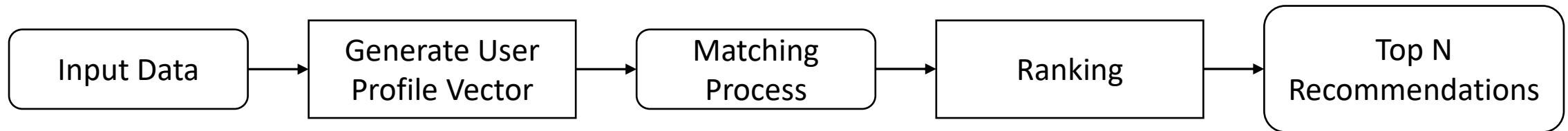


# Top 10 Recommendations (User Profile)

## Top 10 Rekomendasi Kursus (User Profile)

1. DB0101EN
2. ML0101EN
3. PY0101EN
4. DS0103EN
5. CC0101EN
6. BD0145EN
7. ST0101EN
8. WA0101EN
9. CP0101EN
10. DS0101EN

# Flowchart Sistem Rekomendasi Berbasis Konten (Profil Pengguna & Genre Kursus)



--- Logistic Regression ---

Accuracy : 0.3347

RMSE : 1.1769

--- Random Forest ---

Accuracy : 0.3346

RMSE : 1.1509

# Hasil & Evaluasi: Sistem Rekomendasi Berbasis Konten (Profil Pengguna & Genre)

- **Course A** (Genre: Data Science, Python) - Similarity: 0.98
- **Course B** (Genre: Data Science) - Similarity: 0.95
- **Course C** (Genre: Machine Learning, Python) - Similarity: 0.89

--- Logistic Regression ---

Accuracy : 0.3347

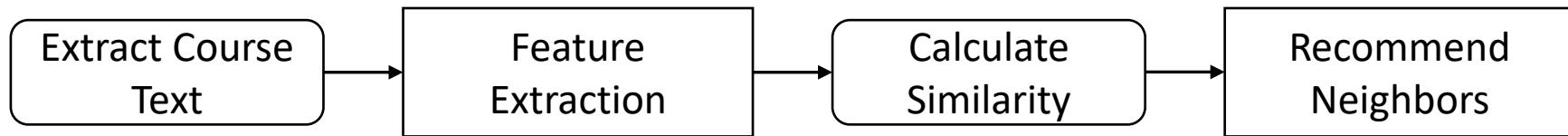
RMSE : 1.1769

--- Random Forest ---

Accuracy : 0.3346

RMSE : 1.1509

# Flowchart Sistem Rekomendasi Berbasis Konten (Kesamaan Kursus)



# Evaluation Results (Course Similarity)

Feature Extraction: TF-IDF (Term Frequency-Inverse Document Frequency)

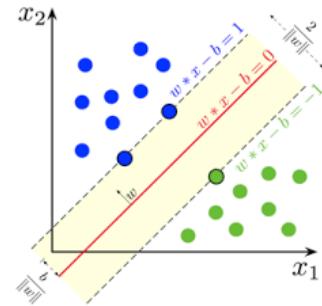
Similarity Metric: Cosine Similarity

Top 10 Courses:

## Top 10 Rekomendasi Kursus (User Profile)

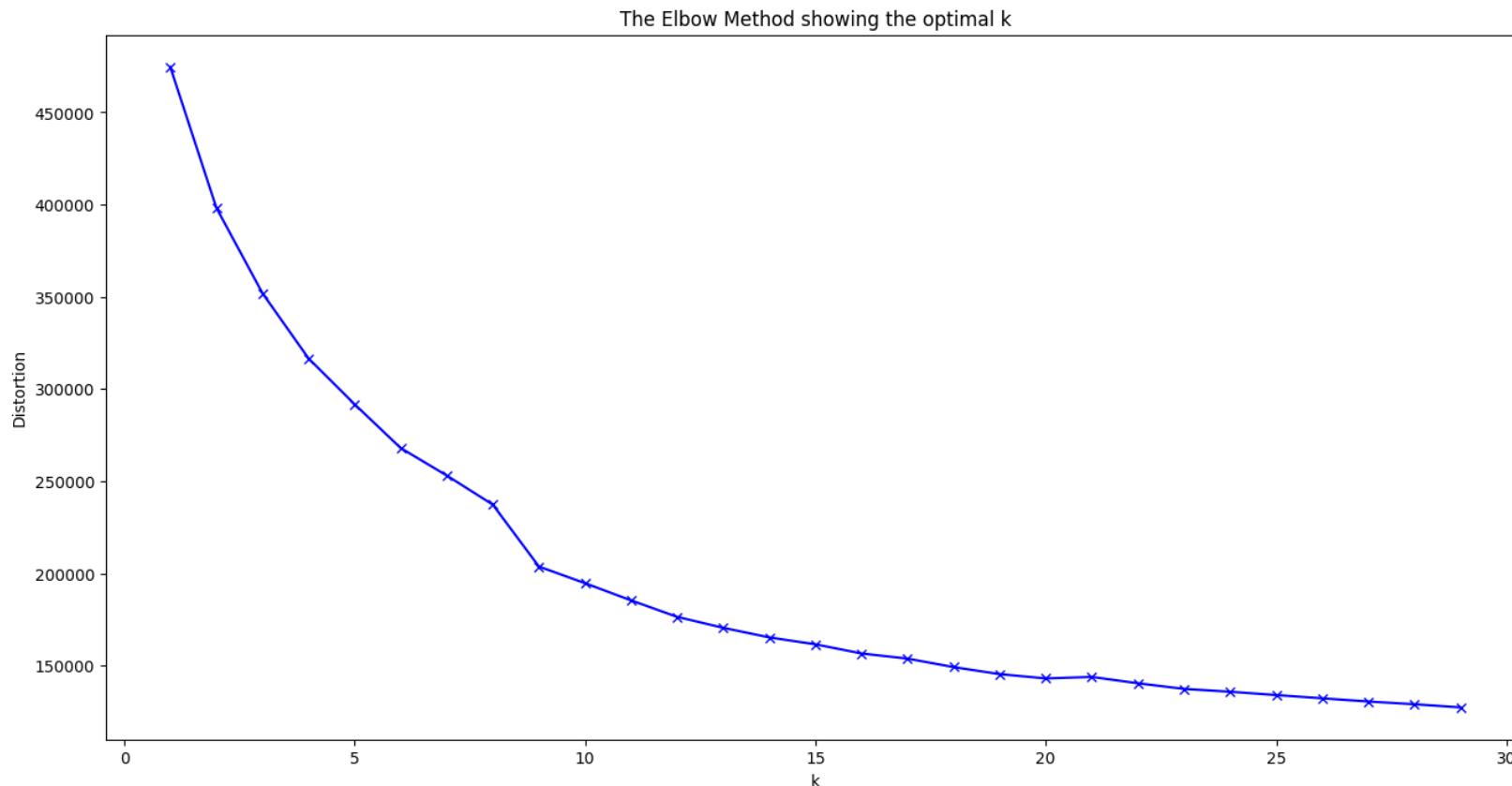
1. DB0101EN
2. ML0101EN
3. PY0101EN
4. DS0103EN
5. CC0101EN
6. BD0145EN
7. ST0101EN
8. WA0101EN
9. CP0101EN
10. DS0101EN

# Collaborative-filtering Recommender System using Supervised Learning



# Evaluation Results (Clustering)

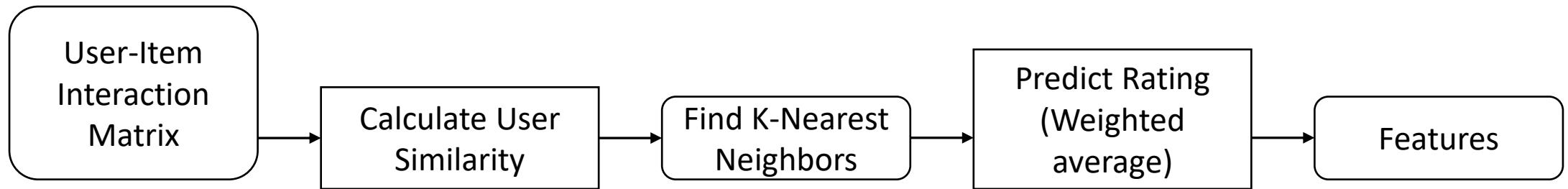
1. Clustering Algorithm: K-Means
2. Dimensionality Reduction: PCA
3. Optimal Clusters ( $k$ ): 14



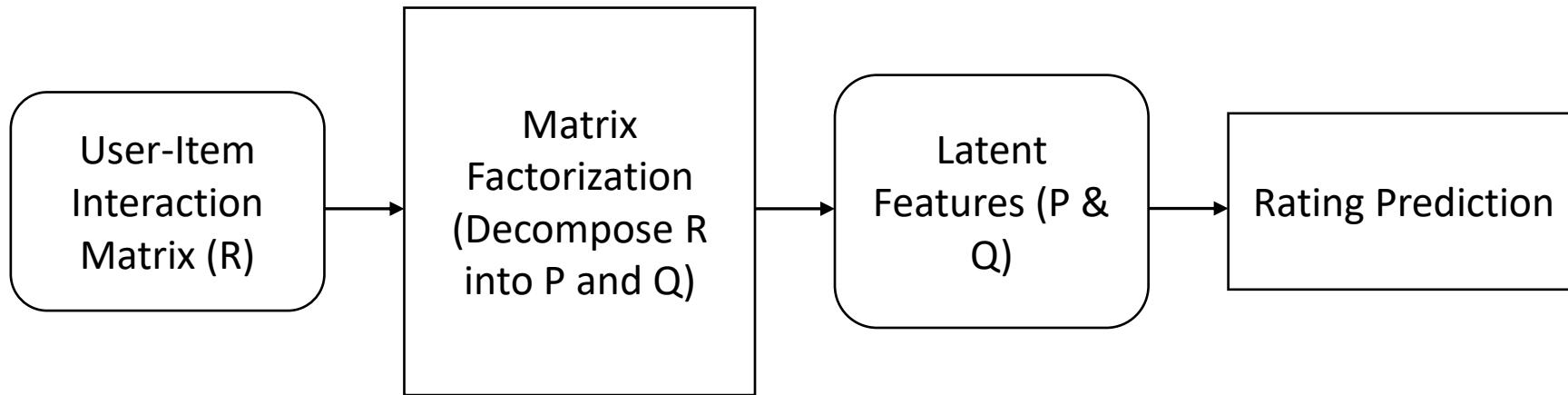
# Clustering Visualization

<b>user</b>	<b>cluster</b>
0	2
1	4
2	5
3	7
4	8
	9

# Flowchart of KNN-based Collaborative Filtering



# Flowchart of NMF-based Collaborative Filtering



# Conclusions

---

- **Model Terbaik:** NMF-based Collaborative Filtering mencapai **RMSE terendah (1.12)**, membuktikan efektivitasnya dalam mengidentifikasi fitur laten di balik interaksi pengguna.
- **Validasi *Embedding*:** Pendekatan Neural Network Embedding menunjukkan performa yang kompetitif (**RMSE 1.15**), memvalidasi penggunaan fitur *deep learning* untuk prediksi.
- **Kekuatan *Content-based*:** Metode *Content-based* (User Profile & Clustering) sangat krusial untuk mengatasi **masalah *cold-start*** (pengguna baru/kursus baru) dengan memberikan rekomendasi yang dapat dijelaskan (*explainable*).
- **Rencana ke Depan:** Melakukan *deployment* model NMF menggunakan *web framework* dan memperluas *feature engineering* untuk memasukkan data temporal (waktu) atau tingkat kesulitan kursus

# Appendix

---

- <https://labs.cognitiveclass.ai/v2/tools/jupyterlab?ulid=ulid-5891f225d69176e55d3e46a1081b511f8b2488fd>
- <https://labs.cognitiveclass.ai/v2/tools/jupyterlab?ulid=ulid-dffd5ec3ec847cd59b139d336a11b3b4fc6da0a>
- **Technology Stack:** Python, Pandas, Scikit-learn, Matplotlib, TensorFlow/Keras (untuk embeddings), Jupyter Notebook.
- **Inovasi & Kreativitas:** Visualisasi PCA yang menunjukkan 14 kluster pengguna yang berbeda dan eksplorasi NMF.