CSC 577 Project Proposal Form

Team name	AI Recommenders
Student names:	Sivaramakrishna Yarra Thusharika Sree Peethala Avinash Betha
Type of project (algorithm-based or evaluation-based) – See Syllabus for descriptions	Algorithm-Based
Brief description of data set:	The Amazon Product Reviews dataset includes user ratings & reviews, product metadata (descriptions, brand, category, price, images), user-item interactions ("also viewed" & "also bought" relationships), and timestamped transactions, offering a comprehensive feature set for personalized recommendations.
Data source:	Amazon Product Reviews dataset from Kaggle.
Recommendation task(s):	• Personalized Product Recommendation: Predict user preferences based on past behavior. • Cold-Start Problem Mitigation: Recommend items using content-based approaches. • Hybrid Model Development: Improve accuracy and diversity by combining multiple techniques.
Algorithm(s): (if algorithm-based, indicate what will be implemented.)	We will develop a hybrid recommender system by integrating multiple techniques. Collaborative Filtering (CF) will include user-based and item-based approaches. Content-Based Filtering (CBF) will leverage TF-IDF, Word2Vec, and product embeddings to extract meaningful features. Matrix factorization methods such as Singular Value Decomposition (SVD) and Alternating Least Squares (ALS) will enhance recommendations. Additionally, deep learning models like Neural Collaborative Filtering (NCF), autoencoders for user and item embeddings, and transformer-based models will further improve recommendation accuracy.
Methodology (description of any experiments to be performed):	This Project will follow a structured methodology: • Data Preprocessing: Clean data, handle missing values, tokenize text (TF-IDF, Word2Vec), and normalize numerical features.

	• Exploratory Data Analysis (EDA): Visualize rating distributions, user-item interactions, and analyze behavioral trends. • Recommendation Models: Implement collaborative filtering (user-based, item-based), content-based filtering (TF-IDF, embeddings), matrix factorization (SVD, ALS), and deep learning models (NCF, autoencoders, transformers). • Hybrid Model Development: Combine CF, CBF, and deep learning models using ensemble methods and meta-learning for optimization. • Model Evaluation & Tuning: Validate with train-test split and optimize using metrics like Precision, Recall, RMSE, MAP, diversity, and novelty. Hyperparameters will be tuned via GridSearchCV or Bayesian Optimization. • Deployment: Build a Flask/Streamlit interface, deploy on AWS/GCP with an API for real-time recommendations, and integrate CI/CD for continuous improvements.
Evaluation metrics to be used:	We will evaluate our recommendation models using Precision, Recall, and F1-Score for accuracy, MSE & RMSE for rating prediction, Hit Rate & MAP for ranking effectiveness, and diversity & novelty metrics to ensure varied recommendations.