



NEXT-GEN TV SHOW RECOMMENDATIONS

DIRECTV

DIRECTV

CHALLENGE ADVISOR

Omer Solakli



Data Scientist,
DIRECTV, Video Technology and Operations
omer.solakli@directv.com



COMPANY OVERVIEW

DirecTV, a leader in sports and entertainment for almost 30 years, provides industry-leading content service and user satisfaction.

- **Mission Statement:** To aggregate, curate, and deliver exceptional, innovative service to customers.
- **Headquarters:** El Segundo, California
- **Employees:** 16,000+
- **Founded:** 1994





MORE ABOUT DIRECTV

- **Subscriber Base:** Over 13 million subscribers
- **Revenue:** \$13 billion in 2023
- **Content Partnerships:** Collaborations with major networks and streaming services
- We have vast library of streamable content from your favorite networks (AMC, ABC, FOX...) on DIRECTV, including TV shows, seasons & episodes.





AI STUDIO CHALLENGE PROJECT OVERVIEW

Challenge Summary: Analyze the IMDB TV Shows dataset to predict show popularity and enhance recommendations using LLM and classical machine learning models.

OBJECTIVE

- Predict TV show popularity based on structured data.
- Analyze the overview column for sentiment and themes using LLM.
- Combine insights to enhance recommendation systems.

DESIRED OUTCOMES

Present predictive models, sentiment analysis results, and actionable insights to improve DirecTV's recommendation engine.

BUSINESS CONTEXT

- Streaming subscription counts continue to climb
- Market is very challenging and demanding of innovations
- Creating the best user experience is essential for a streaming business to survive
- **Project Importance:** Provides insights into viewer preferences, enabling DirecTV to tailor recommendations and improve customer satisfaction.



EXISTING HOMEPAGE

Home Page

LIVE TV YOU MIGHT LIKE



SportsCenter
News | 4:00 – 5:00p



Los Angeles Dodgers at Pittsbu...
Baseball | 3:30 – 6:30p



Summer House
TV Show | 4:00 – 5:00p



The Office
TV Show | 4:45 – 5:20p

POPULAR



The First 48
TV Show | Reality



Tracker
TV Show | Drama



The Chi
TV Show | Drama



Young Sheldon
TV Show | Sitcom

Show Recommendations

- Popular TV Shows
- TV Shows you might like



SUGGESTED APPROACH

- **Problem Type:** Supervised Learning for popularity prediction, NLP for text analysis.
- **Algorithms:**

Popularity Prediction

- Data Preprocessing: Cleaning and encoding data
- Model Training: Using regression models (Linear Regression, Random Forest, Gradient Boosting)
- Model Evaluation: RMSE, MAE, R-squared

LLM Based Theme Trends and Feature Extraction

- Sentiment Analysis: Using pre-trained models like GPT-4 or BERT
- Topic Modeling: Extracting key themes and patterns
- Text Summarization: Generating concise summaries



PROJECT DELIVERABLES

Minimum Requirements

- Analyze IMDB TV Shows dataset to predict show popularity and enhance recommendations
- Give insights on factors influencing TV show popularity
- Common themes in highly rated shows

Key Focus: Predicting popularity based on various show attributes.

Optional Deliverables:

- Analyzing the overview column using Large Language Models (LLMs) for additional insights.
- Trending the common themes by month to forecast which themes are popular or likely to be popular in the future



WORKING WITH LMMS

- Pre-trained Models:** Start with models like GPT-4 or BERT.
- Preprocessing:** Tokenization, normalization of text data.
- Fine-Tuning:** Adapt the model to specific tasks using the IMDB overview data.
- Evaluation:** Use BLEU, ROUGE, and human evaluation.
- Deployment:** Plan for integration and scalability.



DATA OVERVIEW

- **Dataset:** IMDB TV Shows dataset
Key Features: Number of seasons, vote count, vote average, overview, genres, etc.
- **Data Quality:** Ensure completeness and accuracy.
- **Preprocessing:** Handle missing values, normalize numeric data, encode categorical data
- **Data Link:** <https://www.kaggle.com/datasets/asaniczka/full-tmdb-tv-shows-dataset-2023-150k-shows>



PYTHON LIBRARIES

•Common Libraries:

- **Data Manipulation:** pandas, numpy
- **Machine Learning:** scikit-learn, TensorFlow, PyTorch
- **NLP:** NLTK, spaCy, transformers
- **Visualization:** matplotlib, seaborn, plotly





SUGGESTED DEVELOPMENT AND PM TOOLS

- **Development Tools:** Google Colab
- **Collaboration:** GitHub for version control
- **Project Management:** Jira, Notion
- **Frameworks:** Agile, Kanban for task management

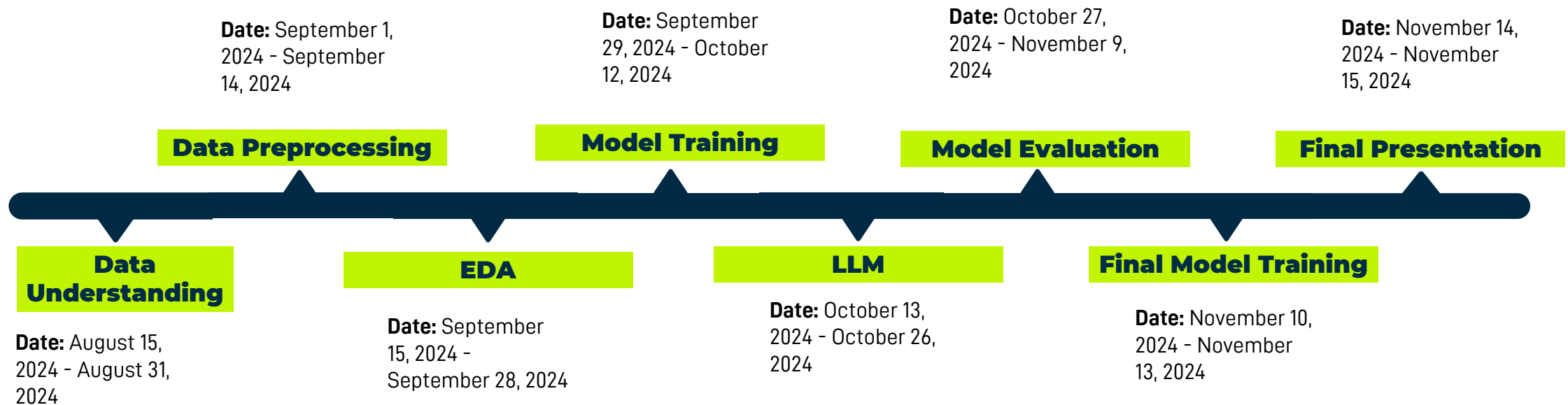


HELPFUL RESOURCES

•Articles and Tutorials

- **Understanding LLMs:** <https://www.youtube.com/watch?v=zizonToFXDs>
- **Regression Analysis:** <https://utsavdesai26.medium.com/linear-regression-made-simple-a-step-by-step-tutorial-fb8e737ea2d9>
- **Data Visualization:** matplotlib, seaborn, plotly

PROJECT MILESTONES AND TIMELINES





MILESTONE 1: DATA UNDERSTANDING AND INITIAL SETUP

Date: August 15, 2024 - August 31, 2024

•Tasks:

- Gather and review the IMDB TV Shows dataset.
- Understand the dataset attributes and initial exploratory data analysis (EDA).
- Set up development environment and tools (Google Colab, GitHub, Notion).





MILESTONE 2: DATA PREPROCESSING

Date: September 1, 2024 - September 14, 2024

•Tasks:

- Clean the dataset (handle missing values, remove duplicates, etc.).
- Encode categorical data and normalize numeric data.
- Preprocess text data in the overview column (tokenization, stop-word removal, etc.).





MILESTONE 3: EXPLORATORY DATA ANALYSIS (EDA)

Date: September 15, 2024 - September 28, 2024

•Tasks:

- Perform detailed EDA to identify key patterns and correlations.
- Visualize data distributions and relationships.
- Generate initial insights and hypotheses for popularity prediction.





MLESTONE 4: MODEL SELECTION AND INITIAL TRAINING

Date: September 29, 2024 - October 12, 2024

•Tasks:

- Select appropriate machine learning models for popularity prediction (Linear Regression, Random Forest, Gradient Boosting).
- Train initial models using preprocessed data.
- Evaluate model performance using metrics (RMSE, MAE, R-squared).





MLESTONE 5: LLM ANALYSIS FOR OVERVIEW COLUMN

Date: October 13, 2024 - October 26, 2024

•Tasks:

- Use pre-trained LLMs to extract embeddings and perform sentiment analysis on the overview text.
- Generate topic models to identify dominant themes.
- Output feature vectors from LLMs to be used in regression analysis.



MILESTONE 6: MODEL EVALUATION AND HYPERPARAMETER TUNING

Date: October 27, 2024 - November 9, 2024

•Tasks:

- Integrate the textual features (embeddings, sentiment, topics) with structured data.
- Train regression models using the combined feature set.
- Evaluate model performance, focusing on the contribution of LLM features.





MLESTONE 7: FINAL MODEL TRAINING AND VALIDATION

Date: November 10, 2024 - November 13, 2024

•Tasks:

- Train the final models using the best configurations.
- Validate the models on a separate test set.
- Ensure robustness and reliability of the predictions.





MILESTONE 8: FINAL PRESENTATION AND REPORT

Date: November 14, 2024 - November 15, 2024

•Tasks:

- Compile results and insights into a comprehensive report.
- Prepare a presentation summarizing the project, methodologies, and findings.
- Present the final outcomes to DirecTV and stakeholders.





HOW WE'LL WORK TOGETHER THIS SEMESTER

- **Check-in Meetings:** Prepare progress updates and questions
- **Reporting:** Weekly updates via email and bi-weekly meetings
- **Communication:** Use company email and Break Through Tech Slack
- **Tools:** Google Colab, GitHub, Notion for project work
- **Project Norms:** Follow Agile methodology with regular sprints



HOW WE'LL WORK TOGETHER THIS SEMESTER

- Check-in Meetings:** Prepare progress updates and questions
- Reporting:** Weekly updates via email and bi-weekly meetings
- Communication:** Use company email and Break Through Tech Slack
- Tools:** Google Colab, GitHub, Notion for project work
- Project Norms:** Follow Agile methodology with regular sprints



HOW TO GET STARTED

Here's what I suggest for your immediate next steps. I'll follow up on your progress and help address any challenges in our next check-in meeting:

Review these slides and note down questions

I'll email you a copy of this deck to store in your Google Drive project folder. Review it as a team and note down any questions you'd like to discuss in our next meeting.

Complete your "Project Scope and Deliverables"

As a team, work on this required Break Through Tech assignment before and during your Sept 7th Maker Day. We'll review it in our next meeting.

Next Steps

Access dataset, set up dev tools, start exploratory data analysis



QUESTIONS?

Prompts

- *What are you most excited about?*
- *Any clarifications needed?*
- *Any initial challenges or concerns?*