HORIZON CONSULTANCY & MARKETING GROUP

PREDICTING HIGH-GROWTH YOUTUBE CHANNELS USING CLASSIFICATION MODEL

BUSINESS UNDERSTANDING

- Horizon Consultancy and Marketing Group is looking forward to partnering with influential YouTubers to help promote several brands that the company will be working with in the next few months.

- The company's Business Development Manager needs information to identify YouTube channels that are likely to experience high subscriber growth in the next month so they can target them for early brand partnerships. Therefore they require to make data-driven decisions based on the predictions from the classification model.

DATA UNDERSTANDING

- The dataset used in this project is Global YouTube Statistics. It offers a perfect avenue to analyze and gain valuable insights from the experts on the platform. It contains 28 columns including:
    - Channel metrics: subscribers, video views, uploads, rank
    - Categorical features: category, channel_type, Country
    - Growth indicators: subscribers_for_last_30_days
    - Country-level context: population, unemployment rate, education enrollment
- Therefore, this makes it suitable for classification. Our target variable derived from this dataset is High growth channel which makes 'subscribers_last_30_days' the binary classification target.

MODELING

- The models use channel-level performance metrics, content characteristics, and country-level socioeconomic indicators as features. These are likely to influence a channel's ability to attract new subscribers.

- A logistic regression model was the baseline model due to its interpretability and effectiveness for binary classification problems

- Secondly, a decision tree was our nonparametric model to capture nonlinear relationships and interactions between various features.

EVALUATION

- The tuned decision tree achieved the highest test accuracy (0.8876), outperforming both the baseline and tuned logistic regression models.

- The strong performance of the decision tree suggests that nonlinear relationships and feature interactions play an important role in predicting high-growth YouTube channels.

- It provides interpretable insights by ranking feature importance, allowing Horizon Consultancy & Marketing Group to better understand which factors most strongly influence channel growth hence will be useful in marketing strategies.

LIMITATIONS

- There are several limitations to be considered when interpreting and applying our model's results.

- First, the model relies on historical and aggregated channel-level data, which may not fully capture short-term trends, content quality, or external factors such as viral events, platform algorithm changes, or shifts in audience behaviour. As a result, predictions may be less accurate for channels experiencing sudden or atypical growth patterns.

- Second, changes in input data i.e small variations in feature values or the introduction of new channels with different characteristics may affect model stability and prediction reliability.

- Finally, some features used in the model, such as demographic or economic indicators, are proxy variables that may not directly influence channel growth but instead correlate with broader market conditions. This introduces uncertainty in causal interpretation and limits the model's ability to explain why growth occurs, rather than simply predicting when it is likely.

.

BUSINESS RECOMMENDATIONS

- Based on our findings, Horizon Consultancy and Marketing Group can use these predictions to more effectively identify YouTube channels with high growth potential and allocate marketing resources strategically.

- Channels predicted to have high growth should be prioritized for partnership opportunities, promotional campaigns, and early-stage brand collaborations, as these creators are more likely to deliver increasing reach and engagement over time. By focusing on these channels, the agency can maximize return on marketing investment while minimizing spending on low-growth prospects.

- Additionally, insights from feature importance analysis suggest that certain channel and market characteristics are strongly associated with growth. Where feasible, the agency can advise clients to optimize controllable factors—such as content consistency, recent engagement momentum, and audience reach—to improve their likelihood of being classified as high-growth
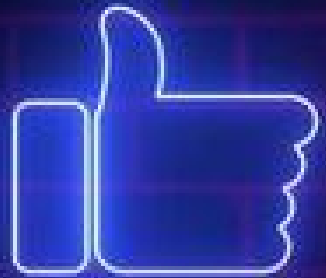
CONT…

- Finally, the model should be used as a decision-support tool rather than a definitive predictor. Combining model outputs with human expertise and qualitative assessments will help ensure that marketing strategies remain flexible and responsive to emerging trends, platform changes, and creative factors not captured in the data.

CONCLUSIONS

- High-growth YouTube channels were identified by building and evaluating multiple models.

- Among the models tested, the tuned decision tree achieved the strongest predictive performance on unseen data, indicating that nonlinear relationships and feature interactions play a significant role in channel growth.

- These findings equip Horizon Consultancy and Marketing Group with a data-driven tool to better target high-growth creators, optimize marketing investments, and make informed strategic decisions while acknowledging the importance of model limitations and contextual use.