Debaterly Market Research

Garrick Kung, Hermes Fu, Josy Pinnick, Matt Russell, Paul Williams, Seojeong Park, Vedanth Dinesh

Introduction

Debaterly offers feature-rich online competitions at fair prices, incorporating cutting-edge technologies like performance analytics and live streaming. Our team's mission is to boost speech & debate competitions, to be as accolade worthy as varsity sports or video games.

Objectives:
- Conduct a competitive and geographical analysis to develop effective pricing strategies and well-structured application models
- Creating a sentiment analysis to differentiate positive and negative speech patterns
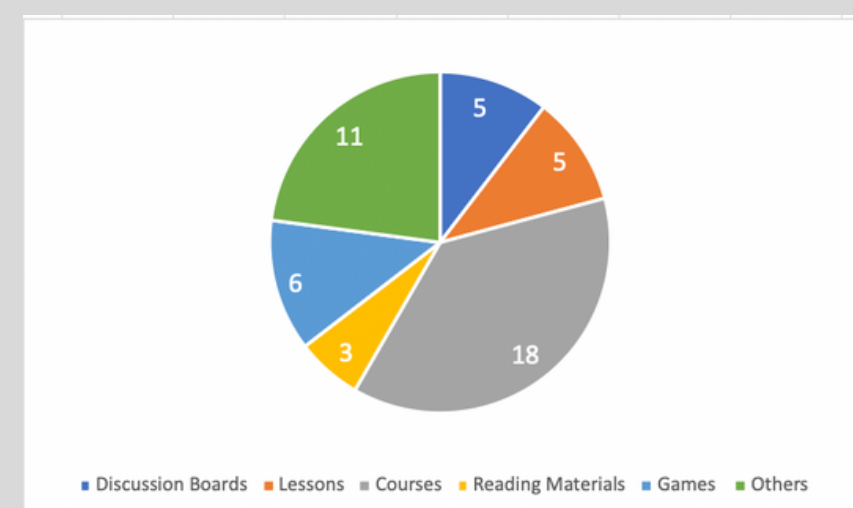
Market Competition & Strategy

Three Trends Observed in the TechEd Sector :
- Education Framework
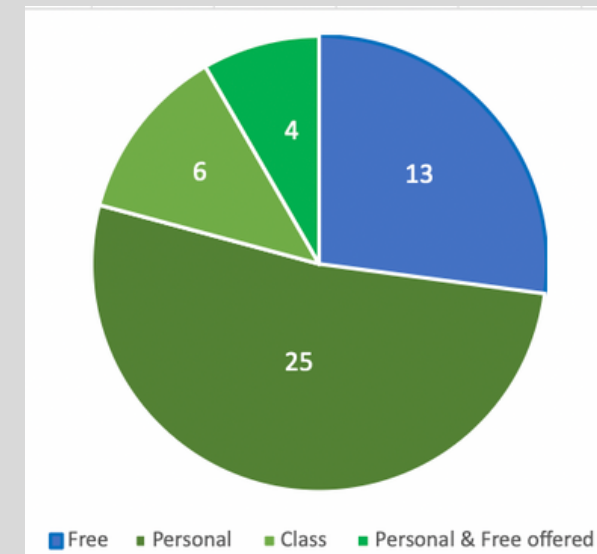- Pricing & Payment Methodology
- Profit Model
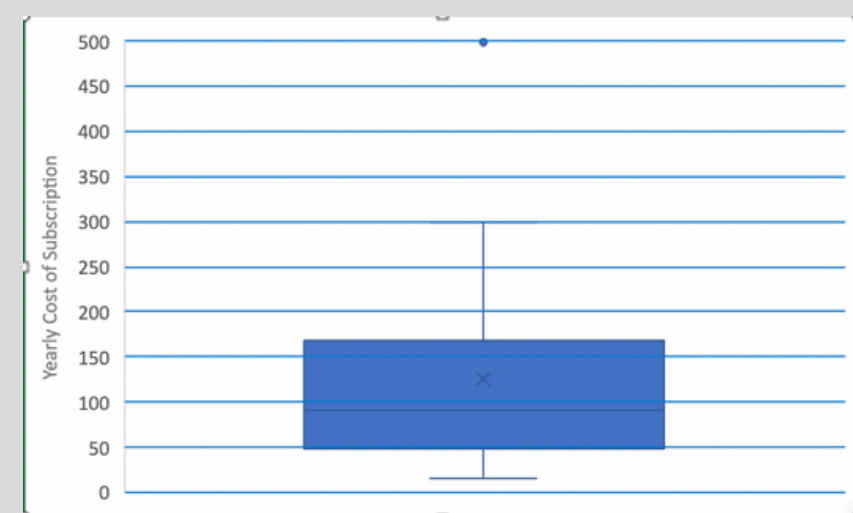
Visualized Subscription Cost of Competitors (Yearly in $)
- Minimum: $20
- Mean: $90
- Maximum: $499

Four Payment Styles Discovered
- Freemium models most prevalent, followed by tiered subscription models

Common Delivery Methods of Education
- Courses most frequent, followed by games, discussion boards, and lessons

Revenue Streams
- Personal plans far more evident than class plans
- Several free options

Future Goals

Debaterly plans to create AI-based learning tools using sentiment analysis. These tools will assist students in the academic markets we have researched. They will grade students' output in real time, provide feedback, etc. The company plans to continue to work with Purdue and the Data Mine. Through this collaboration, Debaterly will progress its deveopment phase and come ever closer to launching the product. The corporate partners and next year's team will continue to work with code, statistical models, and data analysis to create the tools which will be implemented in various schools.

Geographical Analysis

- Exhibits concentration of events in highly populated states, such as California and New York, with outliers in states like Kansas and Utah
- Observed a general overall correlation of event occurrence and population density
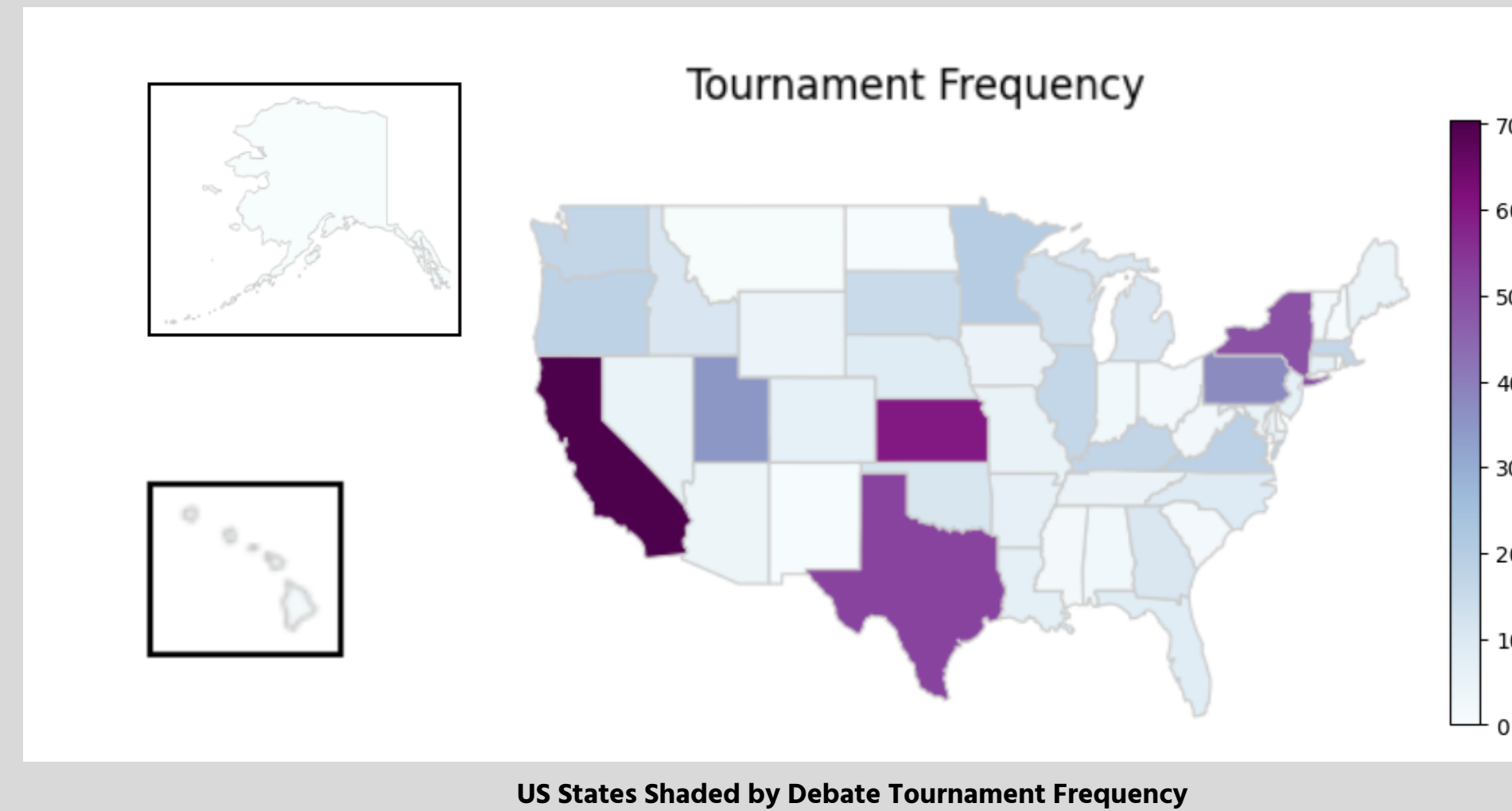
US States Shaded by Debate Tournament Frequency

Debate Competition & Events

Throughout our research on speech & debate competitions around the US, we were able to:
- Facilitate the understanding of the targeted population and informed decision-making
- Create a comprehensive database of past events of importance
- Categorize all teachings into the several types of speech or debate
- Gather demographic insights

Dataset of Debate Tournament Information

Conclusion

Through our team's efforts alongside the Debaterly co-creators, we were able to accomplish each of our objectives and more. Below are some of our major strategic successes this year.
- Identified 48 key competitors and took note of their subscription types, delivery methods of education, and income models
- Deduced the concentration of speech, debate and congressional tournaments across the nation
- Analyzed demographic data in populations of interest to advance toward a smoother eventual launch of the Debaterly application
- Created a sentiment analysis model which achieved over 94.8% accuracy over 20 epochs

Sentiment Analysis

Debaterly wanted to discover what type of debate topic leads to hostile conversations so that Debaterly can prepare and moderate debate topics in a school debate setting to create a safer debate environment for students.

For the sentiment analysis model, we created our custom dataset by collecting debate transcripts and giving them hostility scores. To expand the dataset, we concatenated our custom dataset with a bigger labeled text dataset from Hugging Face to train the model.

- Visualization of current dataset: positive sentiment (larger value) and negative sentiment (lower value) for each transcript
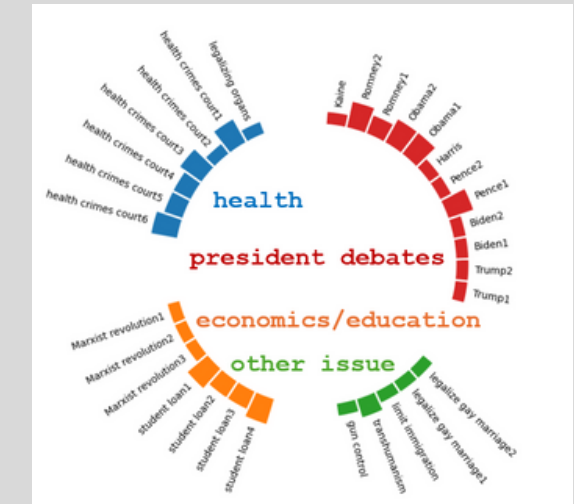
Distribution of Debate Topics

We chose long short-term memory(LSTM) for our sentiment analysis because it is efficient in classifications of variable length and sequential word inputs so that the model can learn long-term dependencies on previous time steps and it also improves the vanishing gradient problem from traditional recurrent neural networks.

Our LSTM model reaches 94.8% accuracy over 20 epochs.

Structure of our LSTM model

Flowchart:
- Concatenate extra large dataset to our custom dataset
- Contraction mapping (e.g., Don't -> Do not) and noise cleaning
- Lemmatization - convert words to their base form
- Pad sequence to ensure all input sequences are of the same length
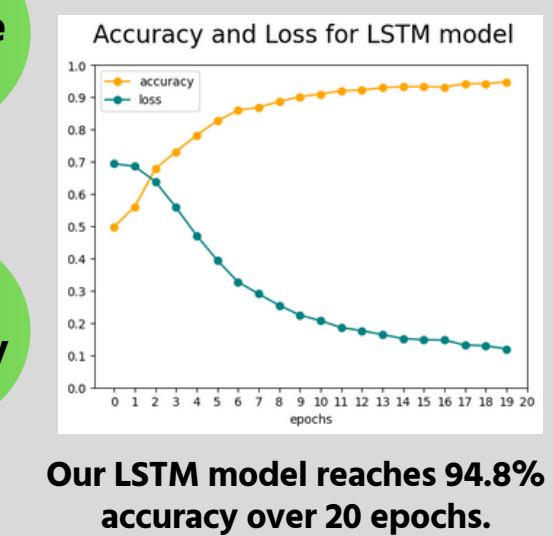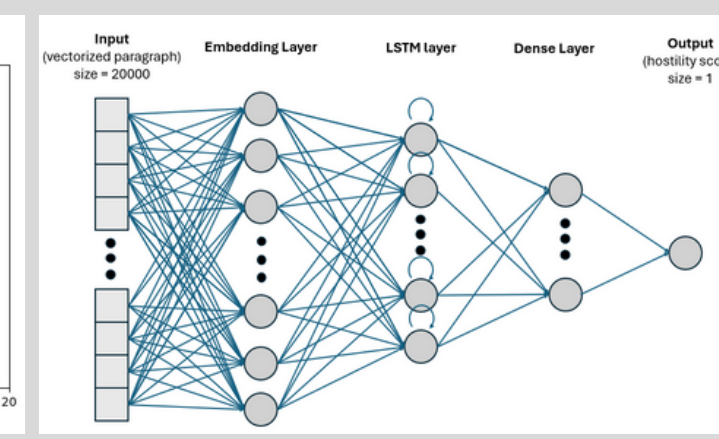- Split data into training and testing set
- Build a sequential neural network with Embedding, LSTM, and Dense layers
- Tweaking hyper-parameters to increase model prediction accuracy

References

- "Index - Tabroom" - https://www.tabroom.com/
- "Tabbycat: Home" - https://calicotab.com/
- "Common Sense Education" - https://www.commonsense.org/education
- "TensorFlow" - https://www.tensorflow.org

Acknowledgements