Clickbait Detector

CS7641-Machine Learning Project Proposal

Narges Moeini, Zachary Aristei, Stephan Tserovski, Jay Cha, Ehsan Asadollahi



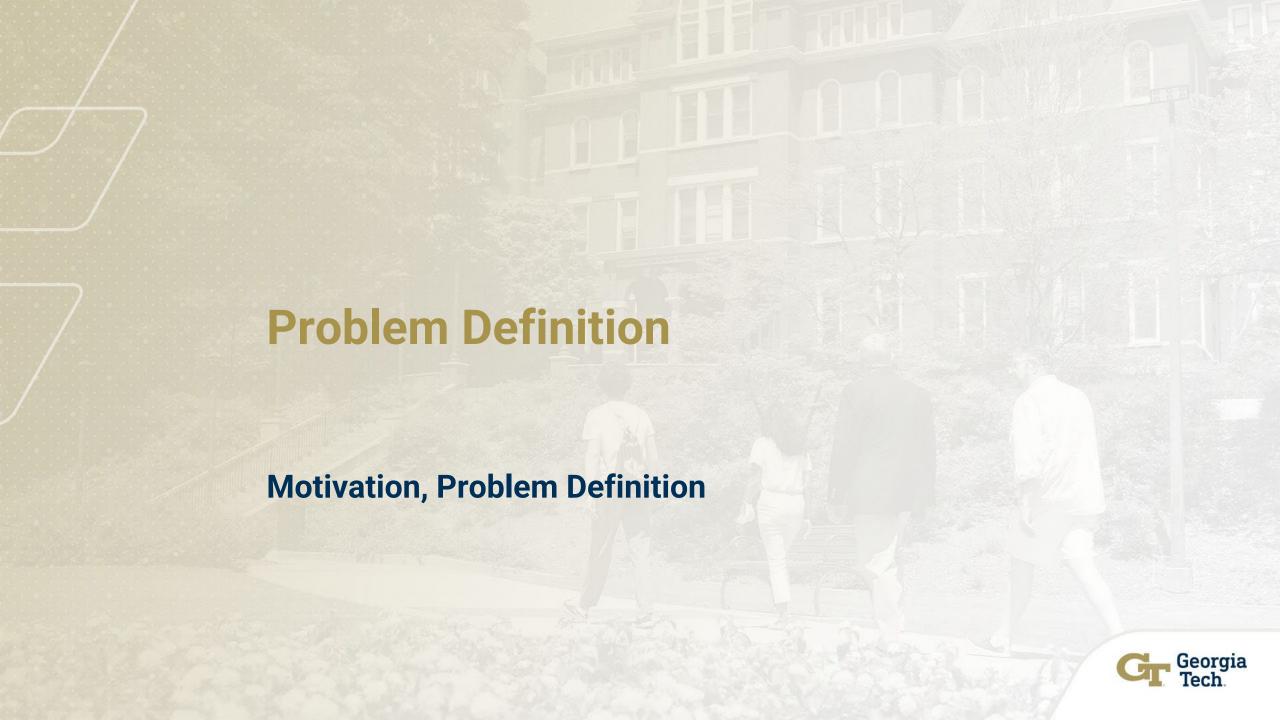


Introduction:

Zachary:

- Content Monetization: Creators use clicks and views to generate revenue.
- Clickbait Issue: Misleading titles drive clicks but frustrate users.





Problem Definition:

Stephan:

- Motivation: Clickbait is far too prevalent in online media
 - Leads to consumers being misled and frustrated
 - Automatically blocking clickbait would lead to a better user experience
- Problem: Make an AI model that can automatically detect clickbait
 - Must be generalized to work across different types of media
 - Other models for clickbait detection tend to have smaller scopes such as only looking at news articles





Methods:

Narges:

- Preprocessing Methods
 - Data Cleaning
 - Feature engineering
 - Data transformation
- ML Methods
 - Unsupervised
 - K-means
 - Supervised
 - Logistic Regression
 - Random Forest





Metrics, expected results, and goals

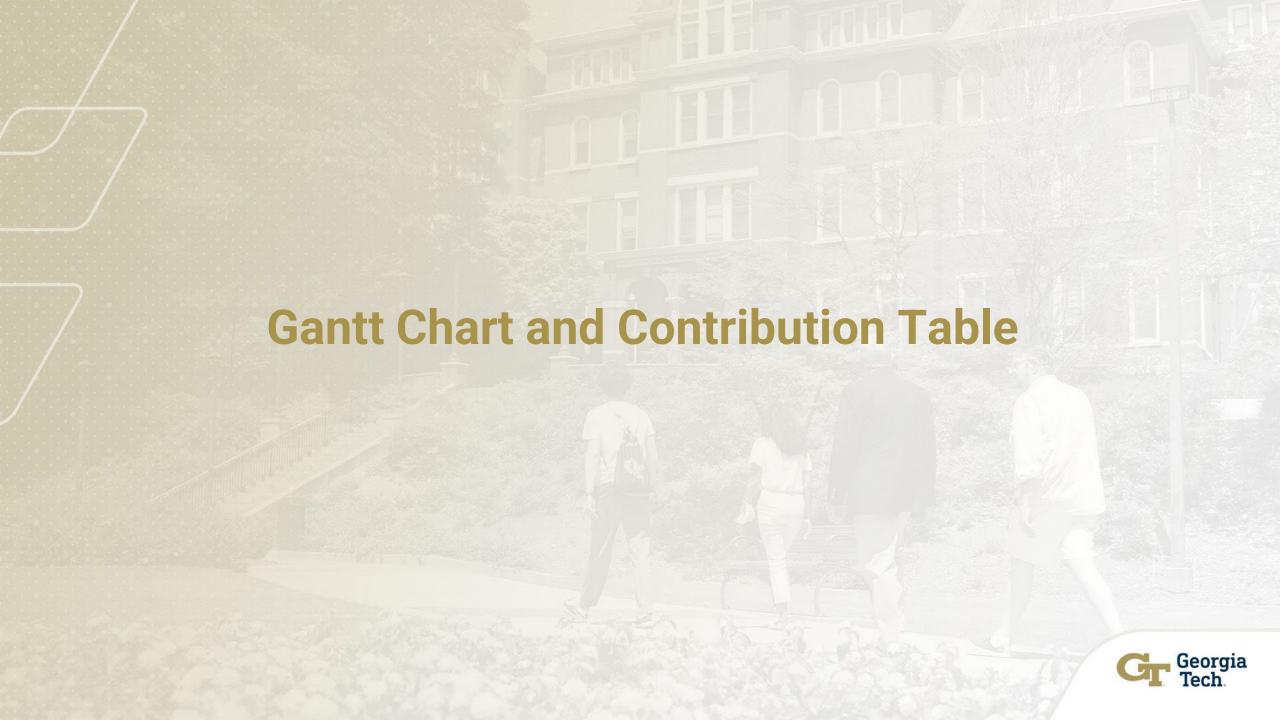


Potential Results and Discussion:

Jay:

- Evaluation metrics
 - Accuracy: Fraction of correct predictions on the test set
 - Precision, recall, F1 score: Information based on true/false positive/negative categories
 - Confusion matrix: Visualize performance across the categories
- Expected results
 - Successfully identify clickbait titles across various media
- Project goals
 - Deploy our model in the real world
 - Allow users to make informed decisions about the content they engage with
 - Discourage the spread of deceptive and harmful practices on the internet





Gantt Chart and Contribution Table:

GANTT CHART

PROJECT TITLE Clickbait Detector												
			PHASE ONE PHASE TWO						PHASE THREE			
		START	Sep 28	Oct 5	Oct 12	Oct 19	Oct 26	Nov 2	Nov 9	Nov16	Nov 23	Nov 30
TASK TITLE	TASK OWNER		S U M T W R	F S U M T W R I	S U M T W R	F S U M T W R F	S U M T W R F	S U M T W R F S	U M T W R F S	U M T W R F S U	M T W R F S U	M T W R F
Project Proposal												
Introduction & Background	Zack	9/28/2024 10/4/2024 7										
Problem Definition	Stephan	9/28/2024 10/4/2024 7										
Methods	Narges	9/28/2024 10/4/2024 7										
Potential Results & Discussion	Jay	9/28/2024 10/4/2024 7										
Video Recording	Ehsan	10/3/2024 10/4/2024 2										
GitHub Page & Streamlit	Ehsan	10/3/2024 10/4/2024 2										
Model 1												
Data Sourcing and Cleaning	Jay	10/5/2024 10/11/2024 7										
Model Selection	Zack	10/5/2024 10/11/2024 7										
Data Pre-Processing	Stephan	10/12/2024 10/18/2024 7										
Model Coding	Ehsan	10/19/2024 11/1/2024 14										
Results Evaluation and Analysis	Narges	11/2/2024 11/8/2024 7										
Midterm Report	All	11/2/2024 11/8/2024 7										
Model 2												
Data Sourcing and Cleaning	Zack	10/12/2024 10/15/2024 4										
Model Selection	Stephan	10/16/2024 10/18/2024 3										
Data Pre-Processing	Ehsan	10/19/2024 10/22/2024 4										
Model Coding	Narges	10/23/2024 11/13/2024 22										
Results Evaluation and Analysis	Jay	11/14/2024 11/19/2024 6										
Model 3												
Data Sourcing and Cleaning	Stephan	10/12/2024 10/15/2024 4										
Model Selection	Ehsan	10/16/2024 10/18/2024 3										
Data Pre-Processing	Narges	10/19/2024 10/22/2024 4										
Model Coding	Jay	10/23/2024 11/13/2024 22										
Results Evaluation and Analysis	Zack	11/14/2024 11/19/2024 6										
Evaluation												
Model Comparison	All	11/25/2024 11/29/2024 5										
Presentation	All	11/25/2024 12/3/2024 9										
Final Report	All	11/25/2024 12/3/2024 9										



Thanks for your attention!

