

XGBoost Predictive Model for TikTok's Claim Classification: EDA, Hypothesis Testing, Logistic Regression, Tree-Based Models

TikTok seeks a predictive model which will reduce the backlog of user reports and prioritize them more efficiently. Machine learning behind the model will extract claims or propositions within TikTok videos and comments and classify each data point as a claim or opinion.

Milestone	Tasks	Deliverables/Reports	Relevant Stakeholder (Optional Activity)
1	<div>Establish structure for project workflow (PACE)</div> <div>Plan ▾</div>	<ul style="list-style-type: none">Global-level project document	Rosie Mae Bradshaw — DS Manager
1a	<div>Write a project proposal</div> <div>Plan ▾</div>		Mary Joanna Rodgers — PM Officer
2	<div>Compile summary information about the data</div> <div>Analyze ▾</div>	<ul style="list-style-type: none">Data files ready for EDA	Orion Rainier — DS
2a	<div>Begin exploring the data</div> <div>Analyze ▾</div>		Orion Rainier — DS
3	<div>Data exploration and cleaning</div> <div>Construct ▾ and Analyze ▾</div>	<ul style="list-style-type: none">EDA report	Orion Rainier — DS

3a	Visualization building Analyze ▾ and Construct ▾	<ul style="list-style-type: none"> Tableau dashboard/visualizations 	Mary Joanna Rodgers — PM Officer
4	Compute descriptive statistics Analyze ▾	<ul style="list-style-type: none"> Analysis of testing results between two important variables 	Willow Jaffey — DS Lead
4a	Conduct hypothesis testing Analyze ▾ and Construct ▾		Orion Rainier — DS
5	Build a regression model Analyze ▾ and Construct ▾		Willow Jaffey — DS Lead
5a	Evaluate the model ▾ Execute ▾	<ul style="list-style-type: none"> Determine the success of the model 	Rosie Mae Bradshaw — DS Manager
6	Build a machine learning model Construct ▾	<ul style="list-style-type: none"> Final model 	Willow Jaffey — DS Lead
6a	Communicate final insights with stakeholders Execute ▾	<ul style="list-style-type: none"> Report to all stakeholders 	Willow Jaffey — DS Lead