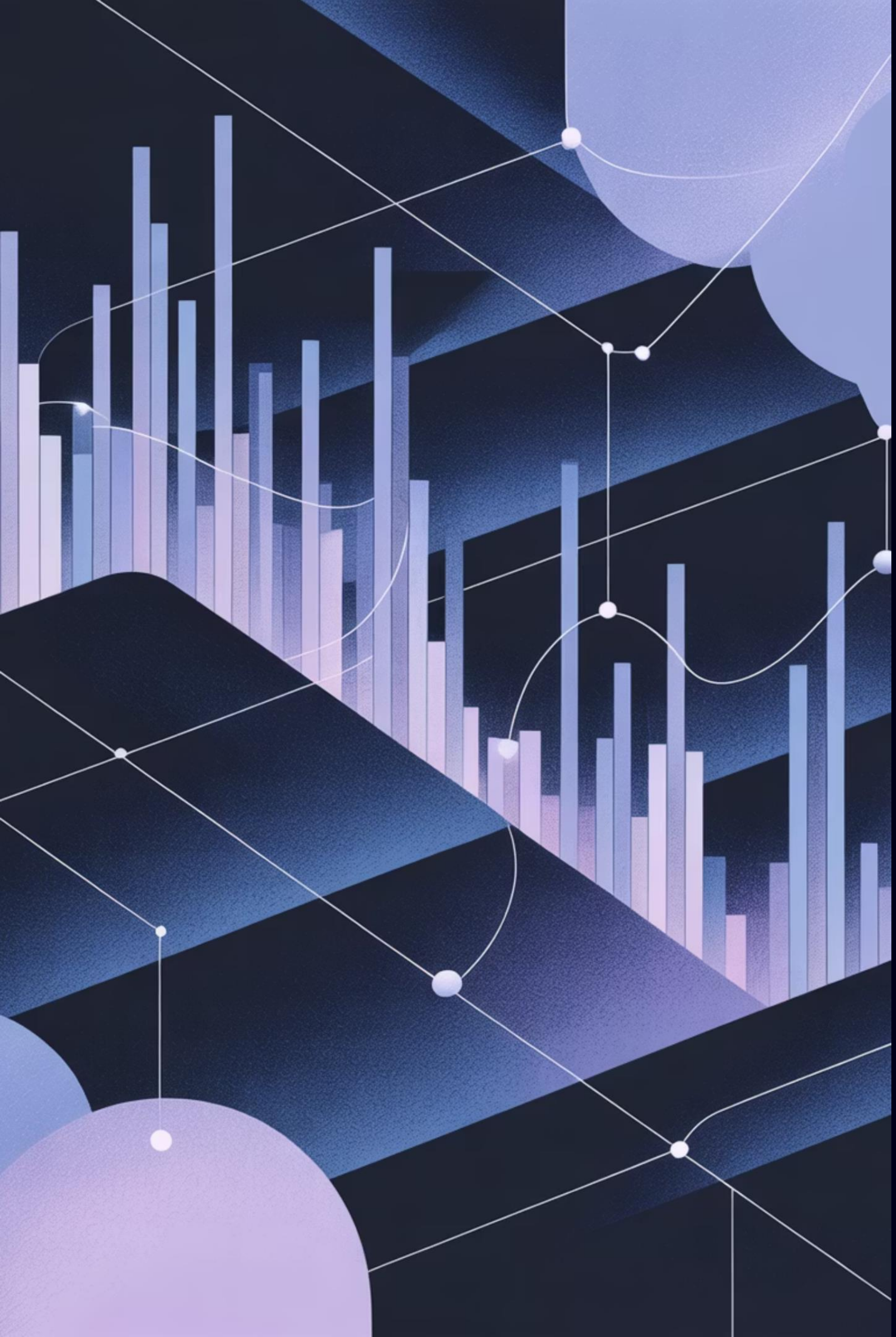


Twitter Sentiment Classification for Consumer Electronics (ElecTech)

CRISP-DM Framework for Sentiment Triage

Unlocking Consumer Insights from Social Media

Social media has become a powerhouse of consumer opinion, particularly for electronics. Businesses are eager to tap into these conversations to inform their strategies.

The challenge lies in the sheer volume and "noise" of tweets. Many are neutral (news, specs) and dilute genuine feedback. We need a way to filter this information effectively.

Our solution is a sentiment classification pipeline. This system will automatically categorise tweets as Positive, Negative, or Neutral, cutting through the noise.

The ultimate goal? Optimised marketing campaigns and smarter inventory decisions, driven by clear, actionable sentiment insights.

The Business Challenge: Sifting Through the Social Media Noise

> Consumer electronics brands are awash in social media data, but the volume of neutral chatter often obscures critical positive and negative feedback. This makes it difficult to extract actionable insights for marketing and inventory management.

Our aim is to overcome this by pinpointing genuine sentiment, enabling a more strategic response to market dynamics.

Project Objectives: What We Aim to Achieve

Mine & Organise Tweets

Extract approximately 9,000 tweets related to Apple and Google products.

Analyse Sentiment

Visualise the distribution of sentiment and uncover initial insights, noting any skew towards neutral or positive categories.

Build Predictive Models

Develop models capable of classifying tweets into Positive, Negative, or Neutral.

Recommend Best Model

Identify and propose the most effective model for deployment and business integration.

Business Understanding: The Foundation of Our Strategy

Understanding consumer sentiment is paramount. It directly impacts marketing effectiveness, product development, and inventory planning.

- **Marketing:** Tailor campaigns to consumer perceptions.
- **Product:** Identify areas for improvement and innovation based on feedback.
- **Inventory:** Adjust stock levels to match anticipated demand influenced by public opinion.

Deliverables:

A clear set of objectives and a detailed project plan for implementing our sentiment triage system.

Data Understanding: Getting to Know Our Tweets

Our dataset comprises approximately 9,000 tweets about Apple and Google products. Each tweet includes vital information to power our analysis:

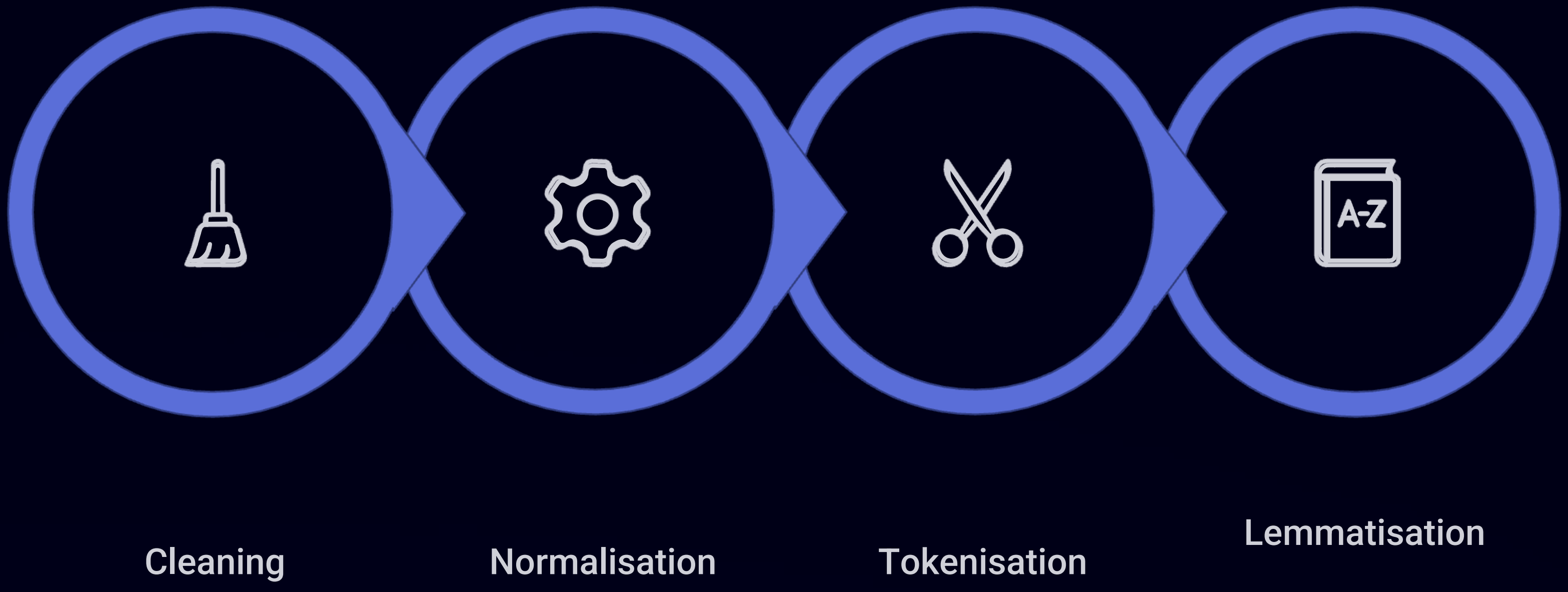
tweet_text	The actual content of the tweet.
sentiment	The assigned sentiment (Positive, Negative, Neutral).
product_brand	Indicates if the tweet is about Apple or Google.

We conducted thorough data quality checks to ensure accuracy:

- **Uniformity:** Ensuring consistent data formats.
- **Null Values:** Identifying and addressing missing data points.
- **Duplicates:** Removing any identical tweets.
- **Ambiguous Labels:** Handling vague sentiment categories like "I can't tell".

Initial analysis reveals a significant prevalence of neutral tweets, highlighting the need for our classification system.

Data Preparation: Refining Raw Data into Actionable Insights



This critical step transforms raw tweet data into a clean, structured format suitable for analysis. We remove noise and standardise text to ensure our models learn effectively.

- **Cleaning:** Removing missing values, duplicates, and ambiguous sentiment labels to ensure data integrity.
- **NLP Preprocessing:** A series of steps to prepare the text:
  - Removing URLs and user mentions.
  - Converting text to lowercase for uniformity.
  - Tokenising text into individual words.
  - Removing common "stop words" (e.g., 'the', 'a', 'is').
  - Lemmatising words to their base form (e.g., 'running' to 'run').

Visualisations confirmed the dominance of neutral tweets, reinforcing the need for precise sentiment classification.

Modeling: Building Our Sentiment Classifier

TF-IDF + Logistic Regression: Our Chosen Foundation

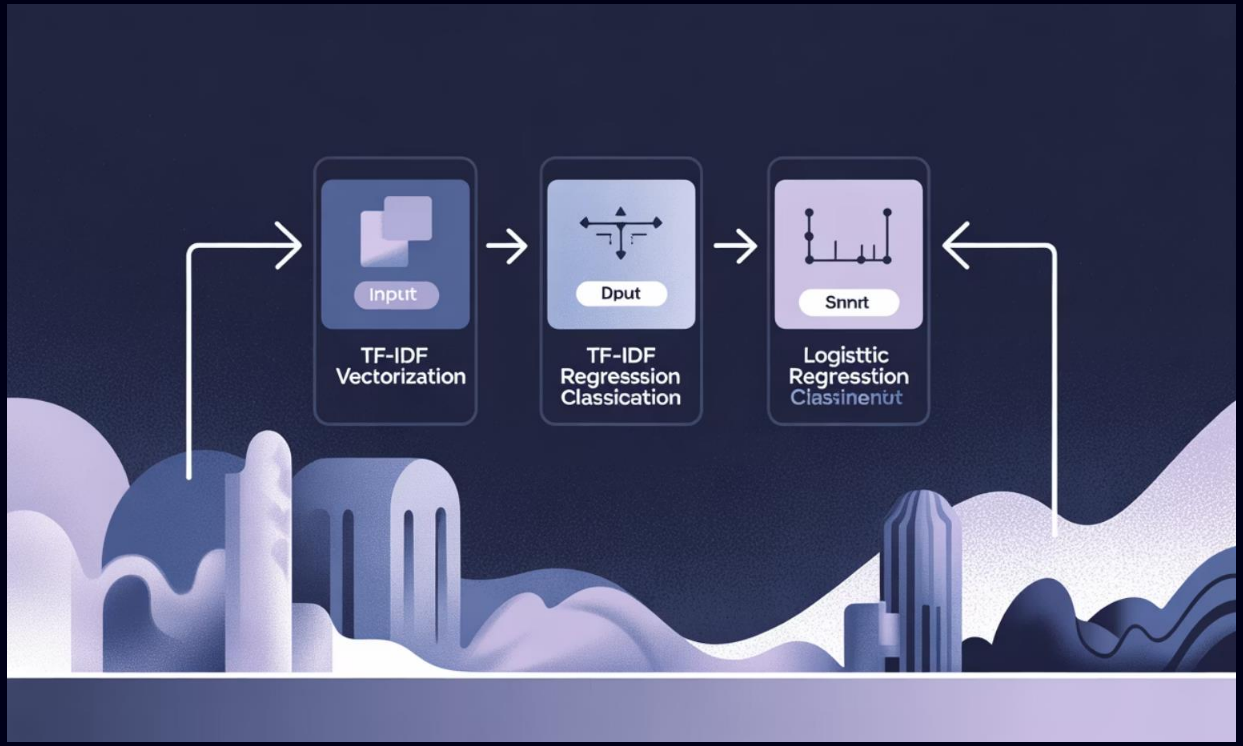
We selected **TF-IDF (Term Frequency-Inverse Document Frequency)** for converting text into numerical features, paired with **Logistic Regression** for classification. This combination offers excellent interpretability and scalability, making it ideal for business application.

Logistic Regression is chosen for its ability to clearly show what drives each sentiment, making it easier for stakeholders to understand the 'why' behind the classifications.

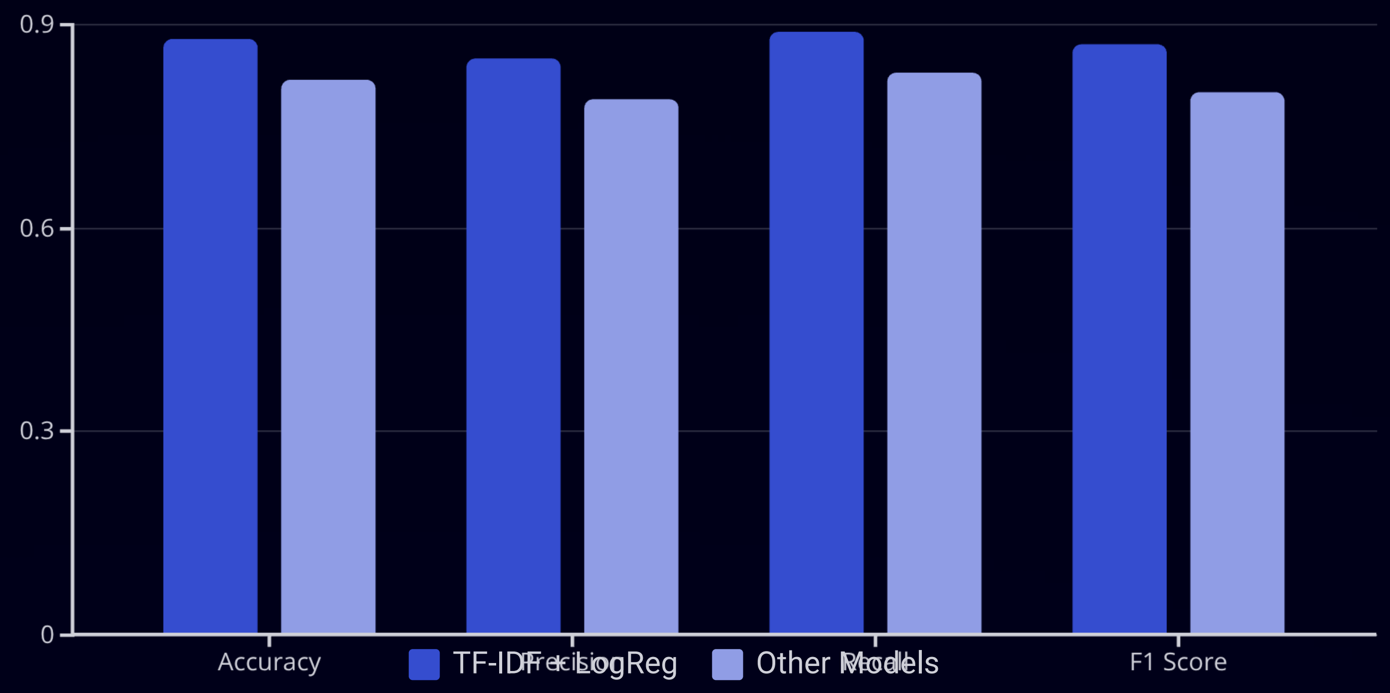
Key Trade-offs:

- **Precision:** Focuses on minimising false positives (fewer irrelevant tweets flagged as important). "Fewer false alarms."
- **Recall:** Focuses on minimising false negatives (fewer important tweets missed). "Fewer missed signals."

Striking the right balance between these ensures we capture essential insights without overwhelming teams with noise.



Evaluation: Measuring Model Performance



We rigorously evaluated our models using key performance metrics, ensuring they meet business requirements:

- **Accuracy:** The overall correctness of predictions.
- **Precision:** How many identified positive/negative tweets are truly positive/negative. (Minimises false alarms).
- **Recall:** How many actual positive/negative tweets were correctly identified. (Minimises missed signals).
- **F1 Score:** A balance between Precision and Recall.
- **AUC Curve:** Measures the model's ability to distinguish between sentiment classes.

Our analysis demonstrates that the TF-IDF + Logistic Regression model consistently outperforms alternatives, striking an optimal balance between performance, interpretability, and business relevance.

Deployment: Bringing the Sentiment Pipeline to Life

Model Storage

The trained model will be saved and stored securely, ready for integration.

Cloud Upload

Uploading the model to a cloud platform ensures accessibility and scalability.

API Creation

An Application Programming Interface (API) will be developed to allow other business systems to easily send tweets and receive sentiment predictions.

User Interface Integration

Integrating the model with existing dashboards or creating a simple user interface will provide real-time sentiment insights to stakeholders.

This deployment strategy ensures the sentiment pipeline is scalable, robust, and directly usable by your teams to make data-driven decisions.