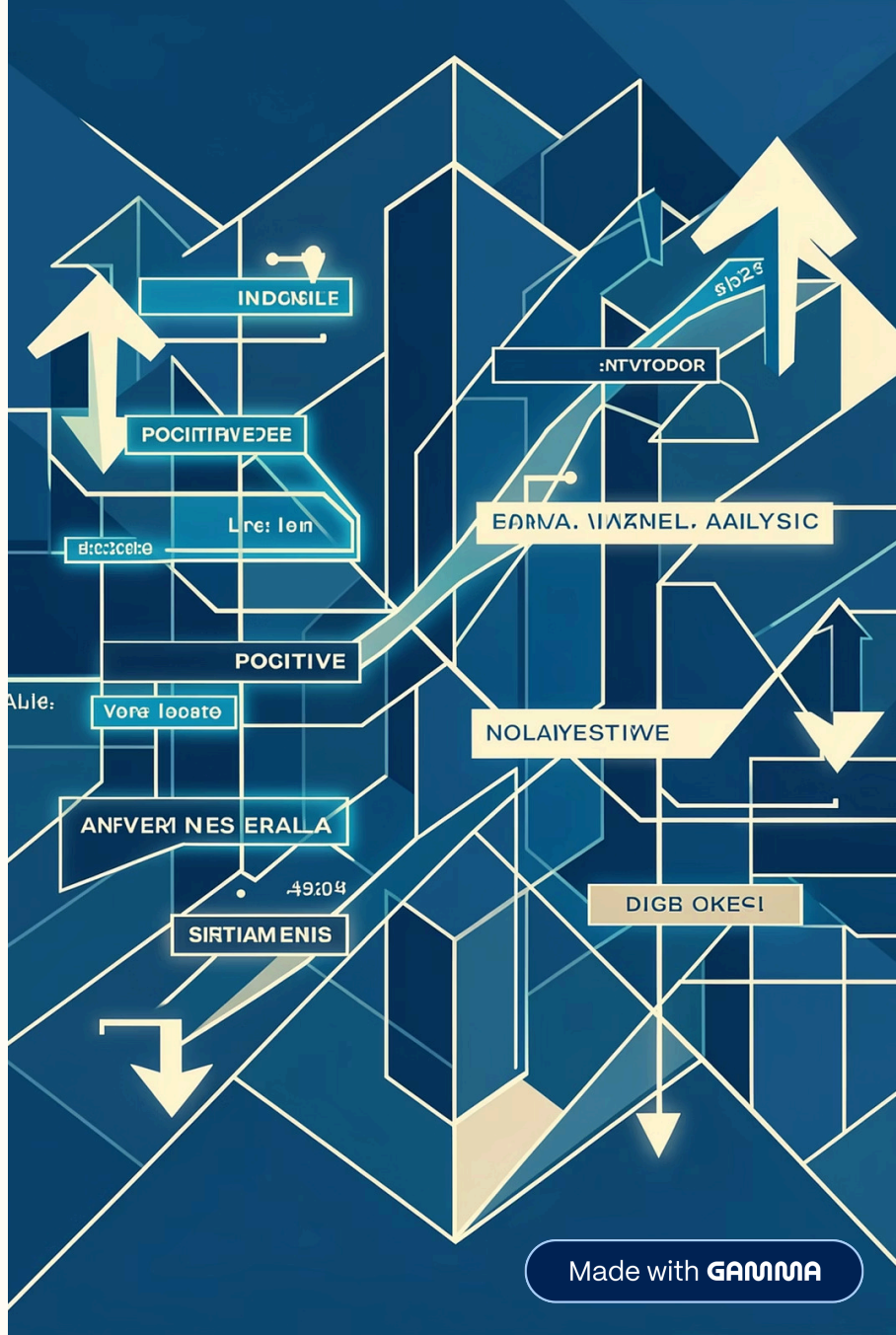
Sentiment Analysis using Machine Learning

Data Analytics Internship

Anisha More



Problem Statement

Analyze text reviews

Classify sentiment (Positive / Negative)

Use NLP techniques

Build ML classification model

Dataset Overview

50,000

Total Records

Text-based review data

Two Sentiment Classes: Positive & Negative

Balanced dataset

Data Preprocessing

01

Converted text to lowercase

02

Removed punctuation & stopwords

03

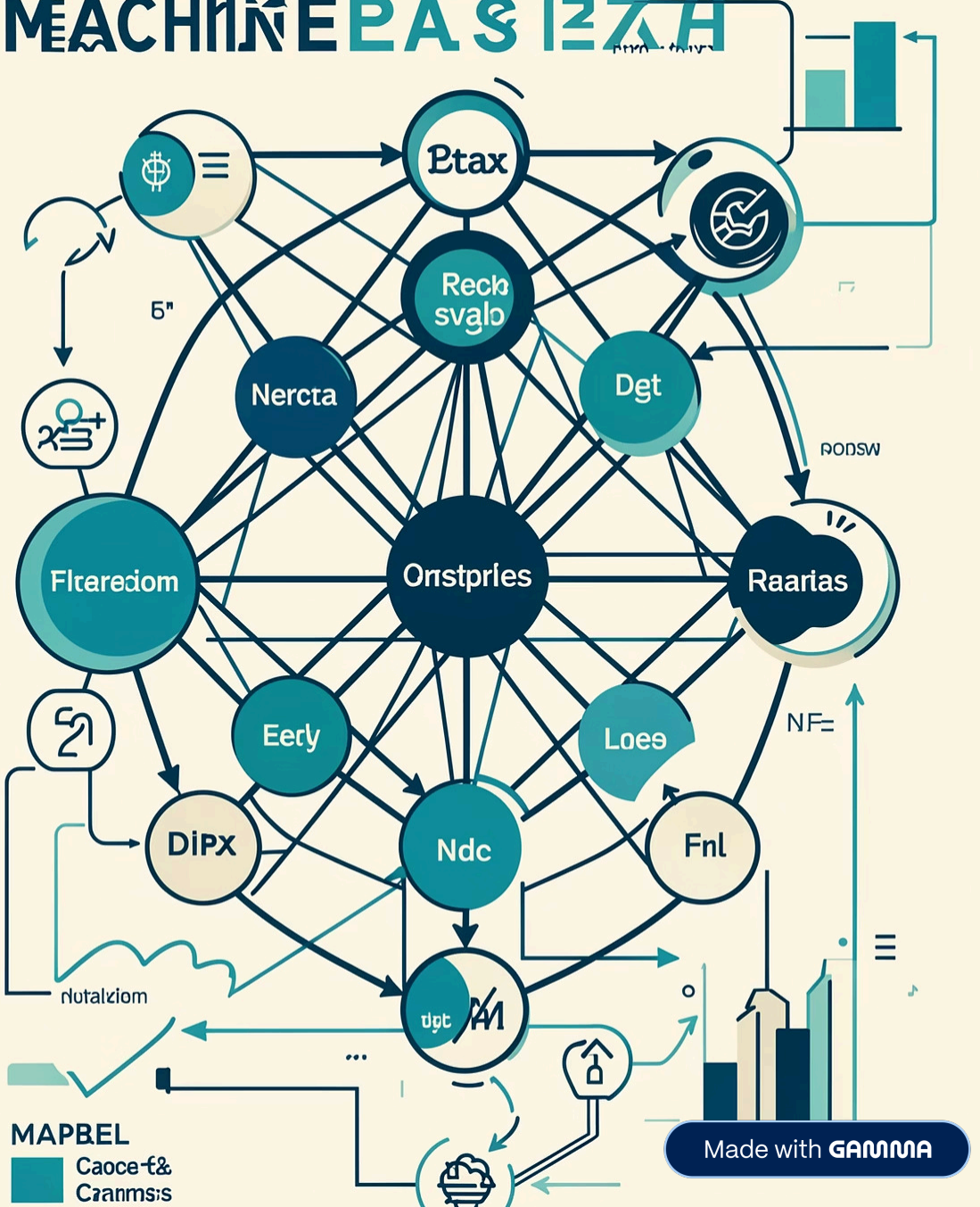
Tokenization

04

Applied TF-IDF Vectorization

Model Used

- Multinomial Naive Bayes
- Train-Test Split (80:20)
- Supervised Learning
- Binary Classification



Model Performance

88%	0.88	0.87	0.87
Accuracy	Precision	Recall	F1 Score

Confusion Matrix Insight

- High true positive rate
- Low misclassification
- Balanced performance

Key Insights

- Model performs well on balanced data
- Positive reviews slightly easier to classify
- Text preprocessing improved accuracy
- TF-IDF effective for feature extraction

Applications

Customer feedback analysis

Social media monitoring

Brand reputation management

Market research

Conclusion

- Successfully built sentiment classifier
- Applied NLP techniques
- Achieved 88% accuracy
- Generated business insights