SOCIAL MEDIA DATA MINING TO GAUGE POSITIVE OR NEGATIVE RESPONSES TOWARDS A COMPANY

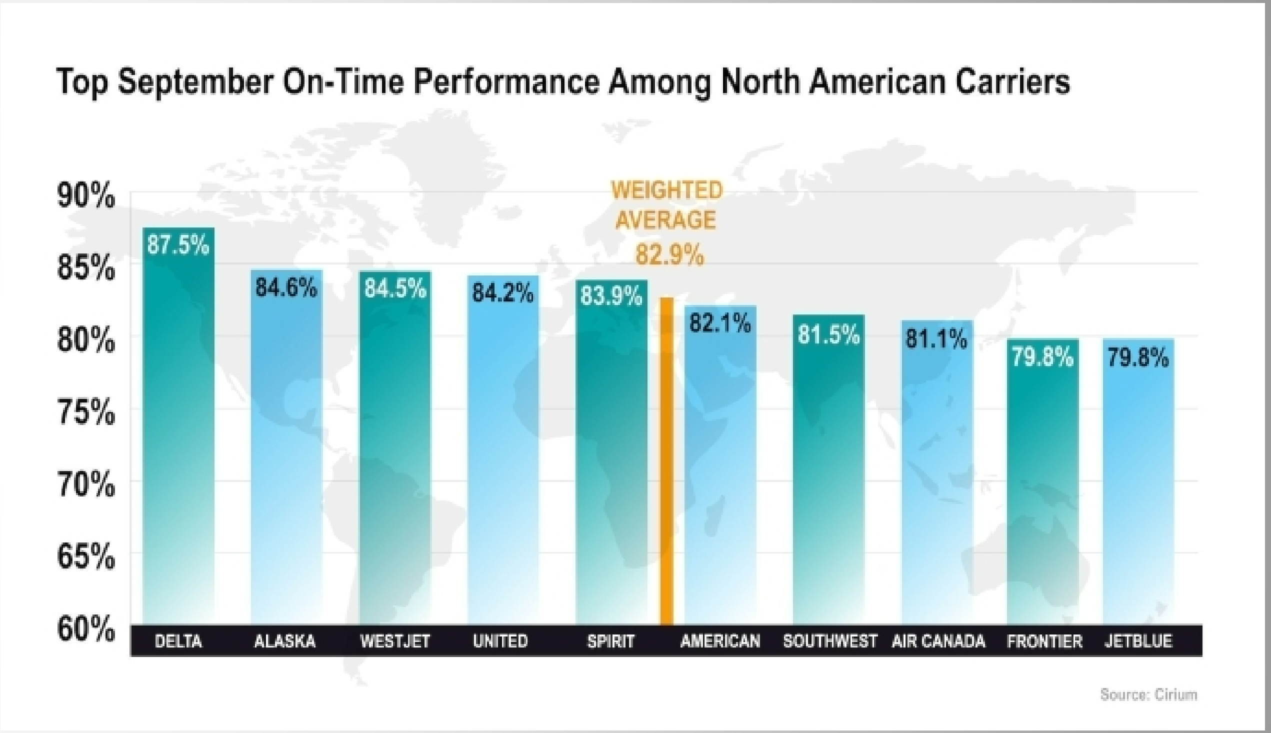
Muhammad Mustafa
CIS 568- Data Mining
University of Michigan Dearborn

PROBLEM OVERVIEW

- Airline passengers post thousands of tweets daily
- Tweets contain real customer experiences
- Unstructured text is hard to analyze manually
- Need automated sentiment classification (Positive / Neutral / Negative)

WHY THIS MATTERS

- Airlines need to understand customer satisfaction trends
- Social media feedback reveals real-time issues
- Data mining helps extract sentiment patterns at scale
- Prior research shows ML and NLP methods perform well on tweet data



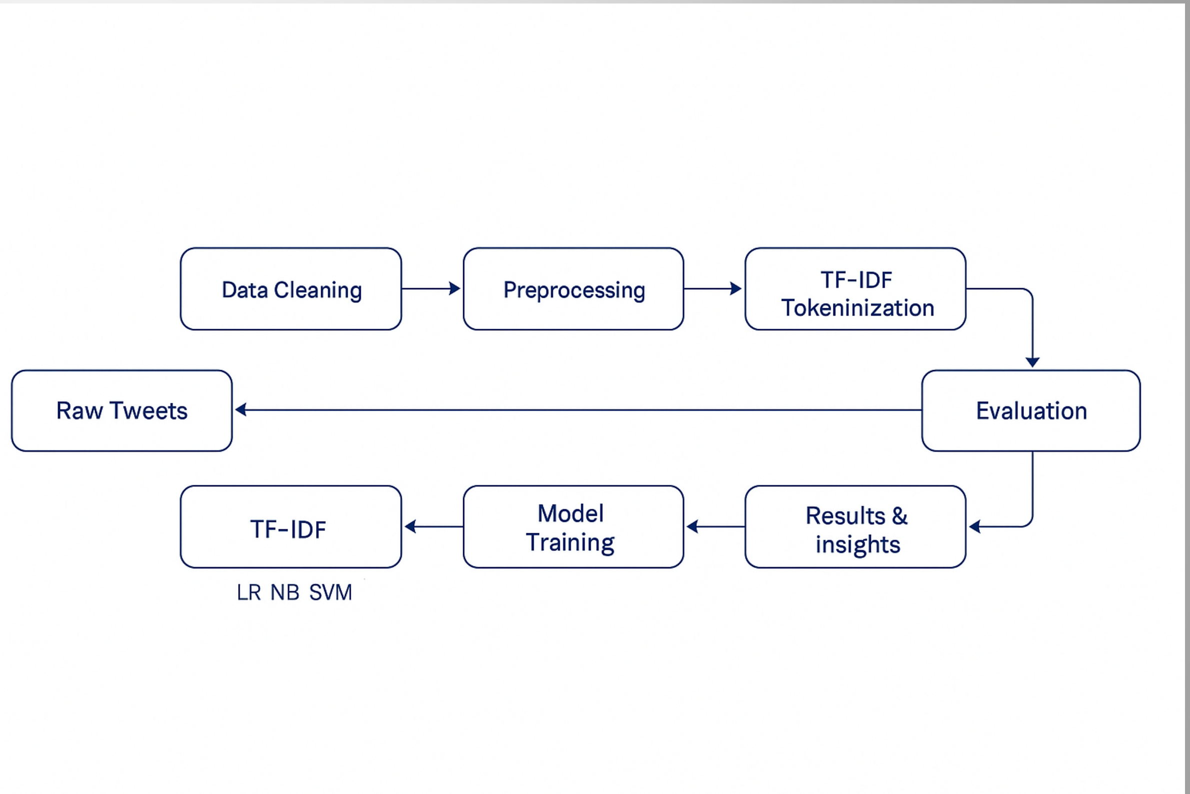
DATASET SUMMARY

- Airline-related tweets with sentiment labels
- Three sentiment classes: Negative, Neutral, Positive
- Key fields used: tweet text, airline, sentiment label, negative reason
- Dataset is imbalanced majority negative
- Provides real-world complexity for model evaluation

airline	airline_se	name	negativere	retweet_cc	text	tweet_coo	tweet_crea	tweet_loca	user_timezone
Virgin America		cairdin		0	@VirginAmerica Wha		########		Eastern Time (US &
Virgin America		jnardino		0	@VirginAmerica plus		########		Pacific Time (US &
Virgin America		yvonnalynn		0	@VirginAmerica I dic		########	Lets Play	Central Time (US &
Virgin America		jnardino		0	@VirginAmerica it's		########		Pacific Time (US &
Virgin America		jnardino		0	@VirginAmerica and		########		Pacific Time (US &
Virgin America		jnardino		0	@VirginA		########		Pacific Time (US &
Virgin America		cjmcginnis		0	@VirginAmerica yes,		########	San Franc	Pacific Time (US &
Virgin America		pilot		0	@VirginAmerica Rea		########	Los Angel	Pacific Time (US &
Virgin America		dhepburn		0	@virginamerica Well		########	San Diego	Pacific Time (US &
Virgin America		YupitsTate		0	@VirginAmerica it w		########	Los Angel	Eastern Time (US &
Virgin America		idk_but_youtube		0	@VirginAmerica did		########	1/1 loner s	Eastern Time (US &
Virgin America		HyperCamiLax		0	@VirginAmerica I &lt		########	NYC	America/New_York
Virgin America		HyperCamiLax		0	@VirginAmerica This		########	NYC	America/New_York
Virgin America		mollanderson		0	@VirginAmerica @vi		########		Eastern Time (US &
Virgin America		sjespers		0	@VirginAmerica Tha		########	San Franc	Pacific Time (US &
Virgin America		smartwatermelon		0	@VirginAmerica SFO		########	palo alto,	Pacific Time (US &
Virgin America		ItzBrianHunty		0	@VirginAmerica So e		########	west covir	Pacific Time (US &
Virgin America		heatherovieda		0	@VirginAmerica I fle		########	this place	Eastern Time (US &
Virgin America		thebrandiray		0	I â¤ï¸ flying @VirginA		########	Somewhe	Atlantic Time (Can
Virgin America		JNLpierce		0	@VirginAmerica you		########	Boston | V	Quito
Virgin America		MISSGJ		0	@VirginAmerica why		########		
Virgin America		DT_Les		0	@VirginAr	[40.748042	########		
Virgin America		ElvinaBeck		0	@VirginAmerica I lov		########	Los Angel	Pacific Time (US &
Virgin America		rjlynch21086		0	@VirginAmerica will		########	Boston, M	Eastern Time (US &
Virgin America		ayeevickiee		0	@VirginAmerica you		########	714	Mountain Time (US
Virgin America		Leora13		0	@VirginAmerica stat		########		
Virgin America		meredithjlynn		0	@VirginAmerica Wha		########		
Virgin America		AdamSinger		0	@VirginAmerica do y		########	San Franc	Central Time (US &
Virgin America		blackjackpro911		0	@VirginAr	[42.361016	########	San Mateo, CA & Las Vegas, N	
Virgin America		TenantsUpstairs		0	@VirginAr	[33.945404	########	Brooklyn	Atlantic Time (Can
Virgin America		jordanpichler		0	@VirginAmerica hi I		########		Vienna

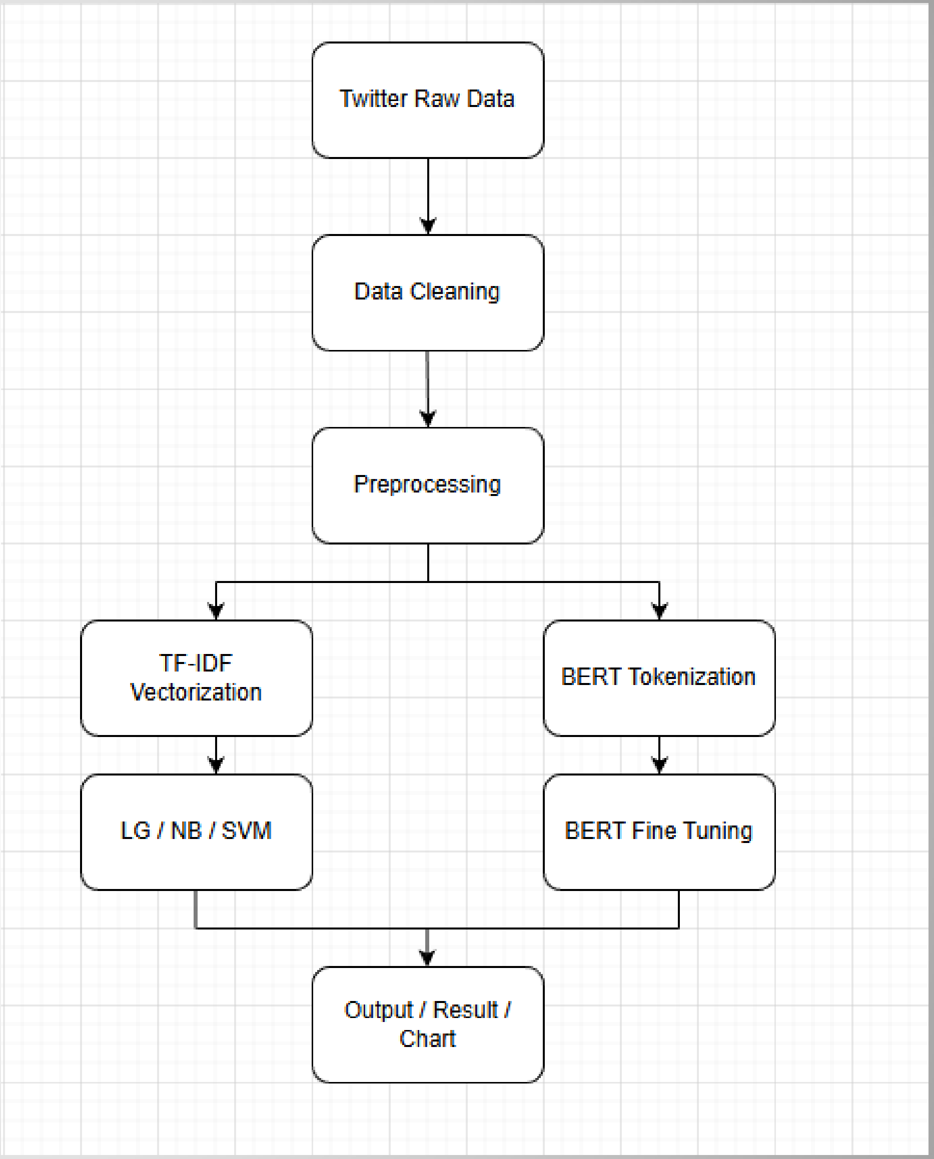
DATA PREPARATION

- Social media text contains noise and inconsistencies
- Cleaning ensures high-quality input for models
- Preprocessing improves model accuracy and stability
- Supports both classical ML and BERT pipelines



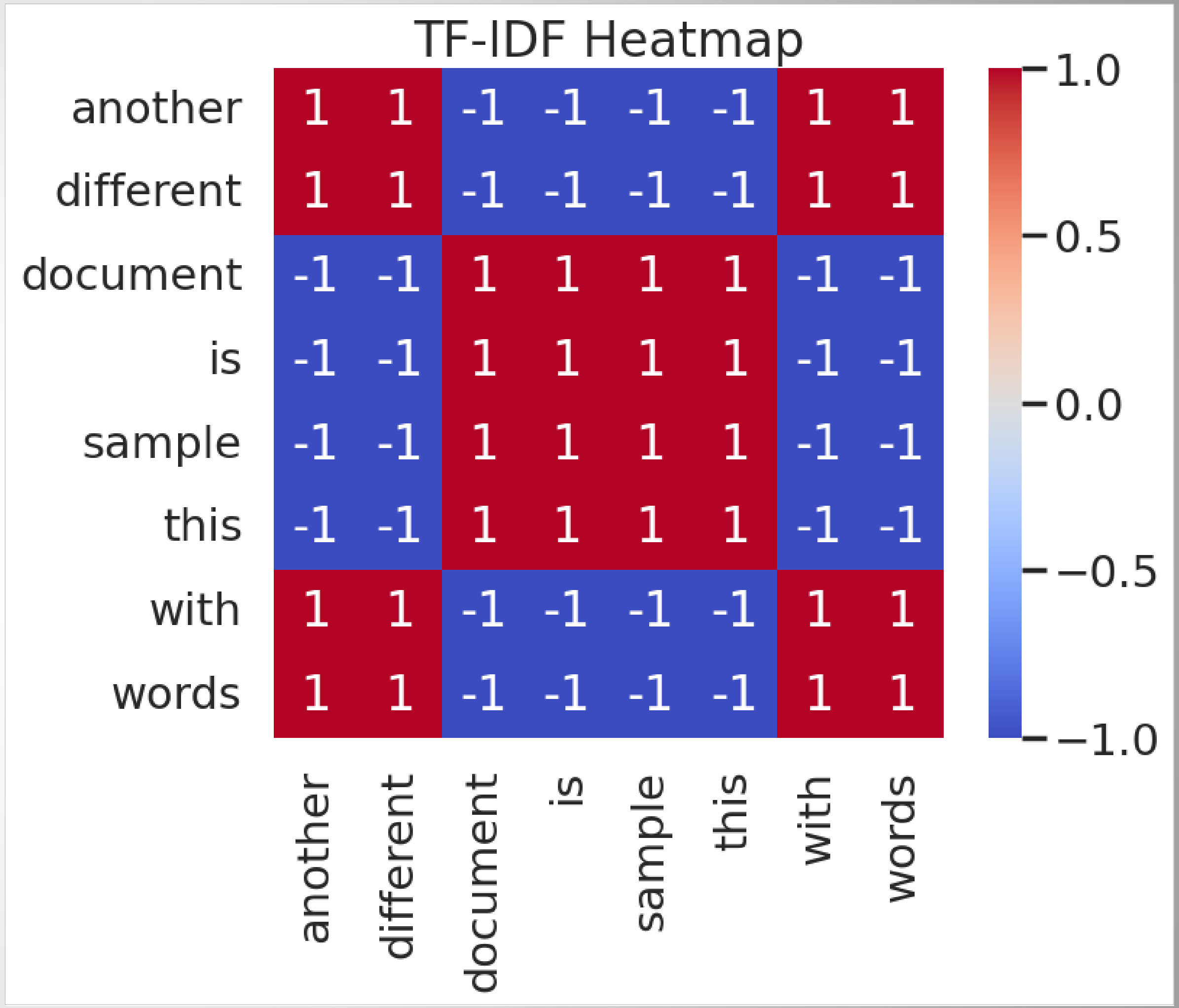
DATA FLOW DIAGRAM

- Both ML and BERT follow shared preprocessing
- TF-IDF powers LR, NB, SVM
- Tokenization powers BERT fine-tuning
- All models evaluated on the same test split



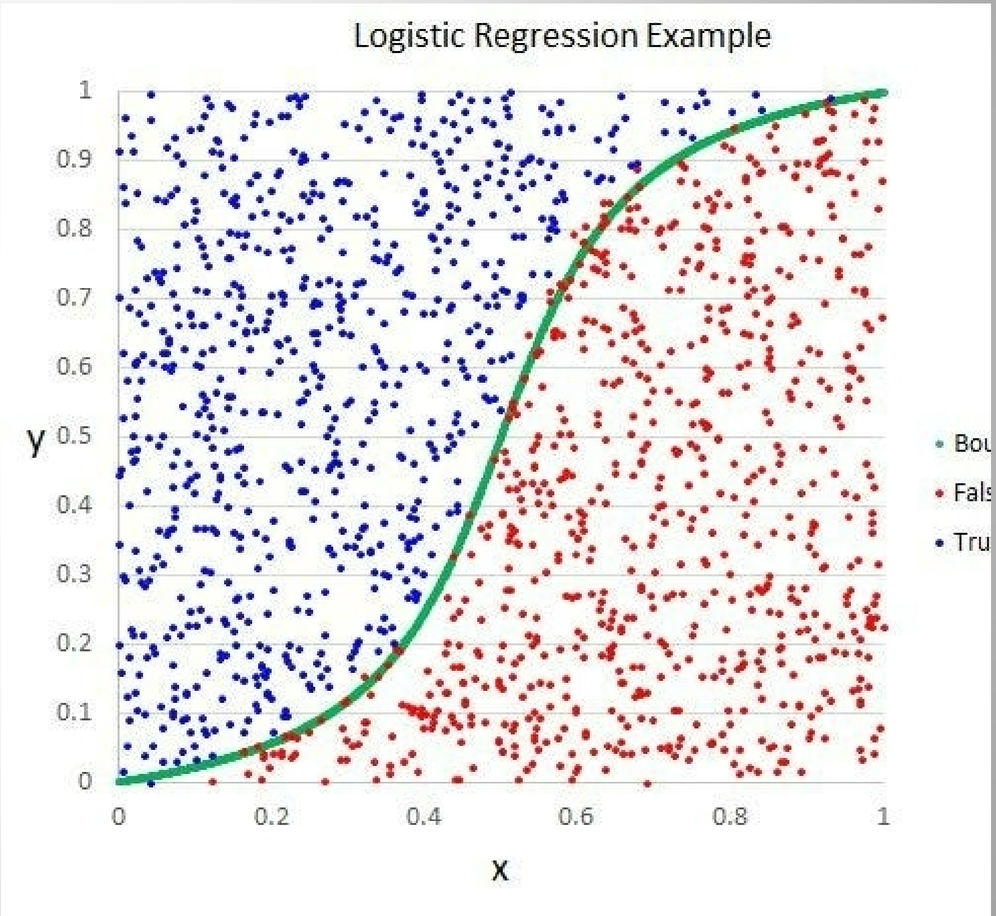
FEATURE EXTRACTION (TF-IDF)

- Converts text into numerical vectors
- Captures important words across tweets
- Works extremely well for classical ML models
- Uses unigrams and bigrams for richer representation



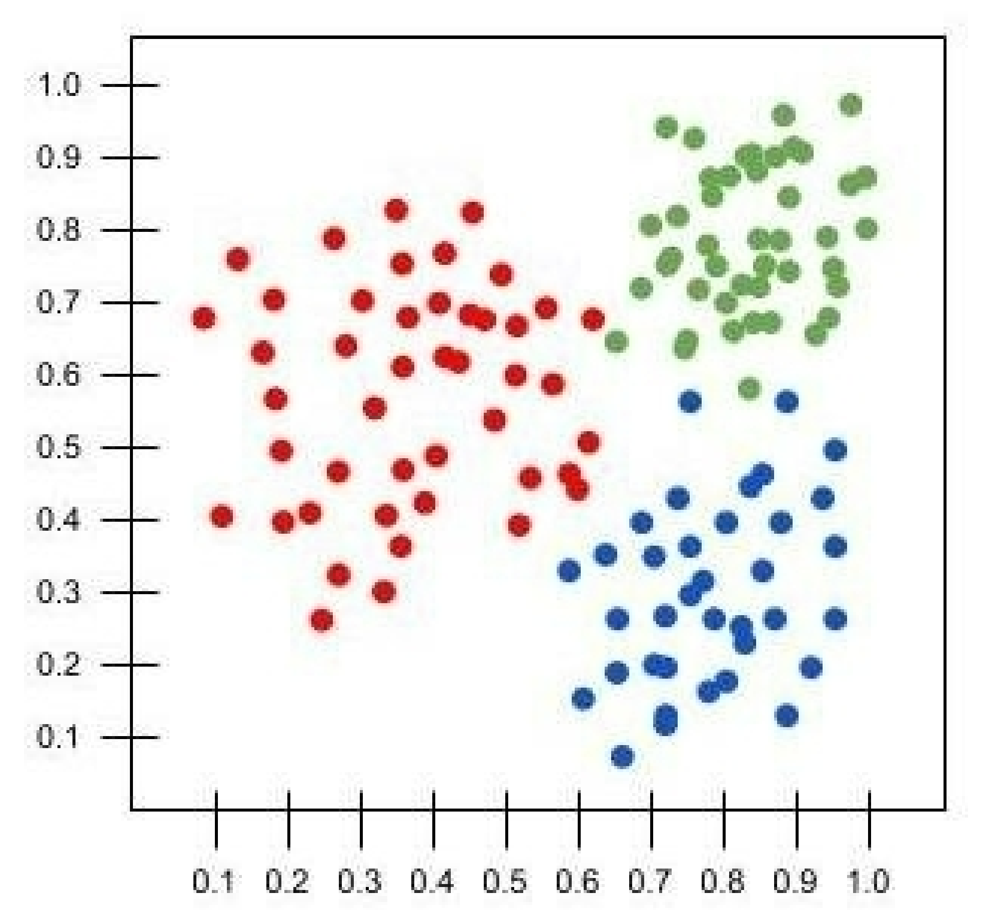
MODELS USED
LOGISTIC REGRESSION

- Works well with high-dimensional sparse text data
- Interpretable weight coefficients show which words influence sentiment
- Fast training + strong baseline performance
- Handles linearly separable classes effectively



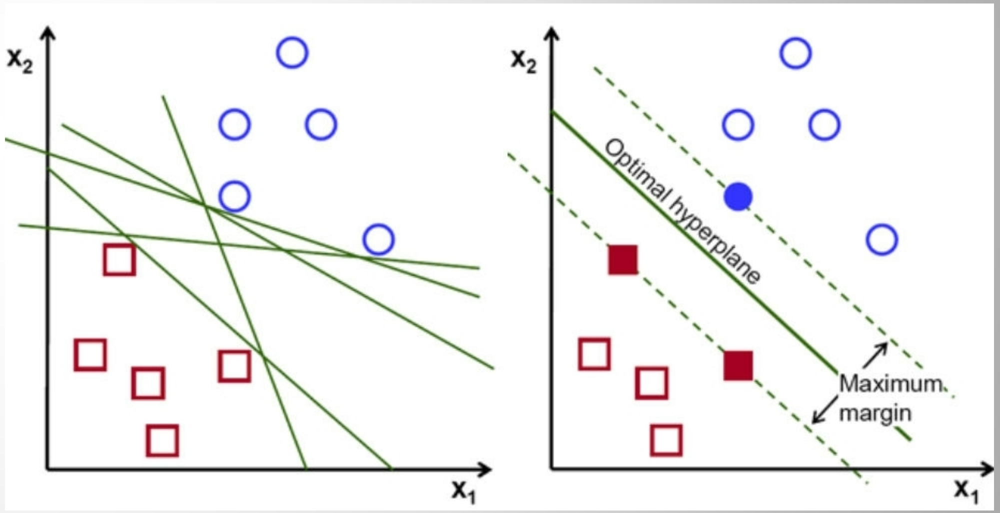
NAIVE BAYES

- Based on probability of words given each sentiment class
- Very fast and efficient for large text datasets
- Performs well when word distributions differ strongly across classes
- Assumes independence between words (can be a limitation)



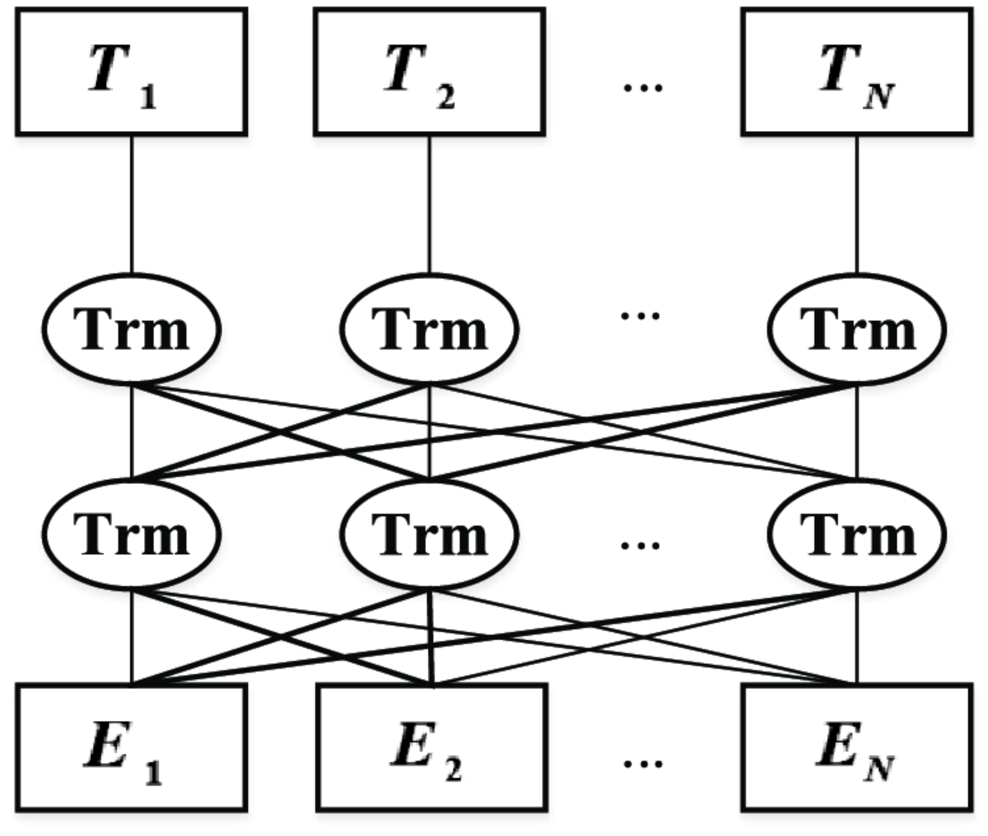
SUPPORT VECTOR MACHINE (SVM)

- Maximizes the margin between sentiment classes
- Very effective with TF-IDF vectors
- Handles high-dimensional text data better than many models
- Typically produces the highest accuracy in classical text classification tasks



BERT (BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS)

- Reads entire sentence context (before + after each word)
- Generates powerful contextual embeddings for each token
- Pretrained on massive corpus strong language understanding
- Fine-tuned on airline sentiment for better task-specific accuracy
- Handles sarcasm, negation, and nuanced tweet language better than TF-IDF models
- More computationally expensive (GPU recommended)



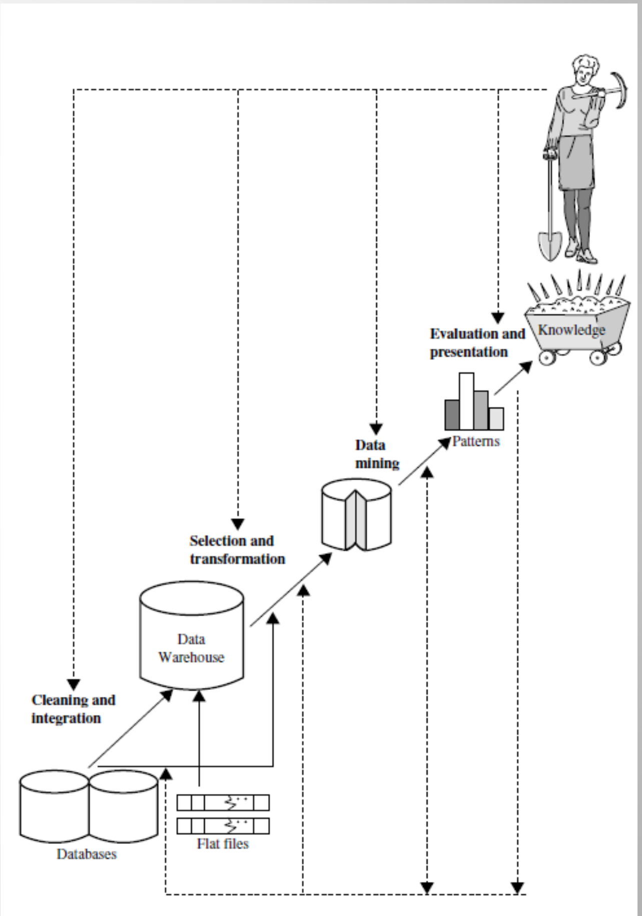
EXPERIMENTAL SETUP

- 80/20 train-test split
- TF-IDF features - LR, NB, SVM
- Token IDs + attention masks - BERT

Evaluation metrics:

- ➢ Accuracy
- ➢ Precision, Recall
- ➢ F1-Score
- ➢ Confusion Matrix

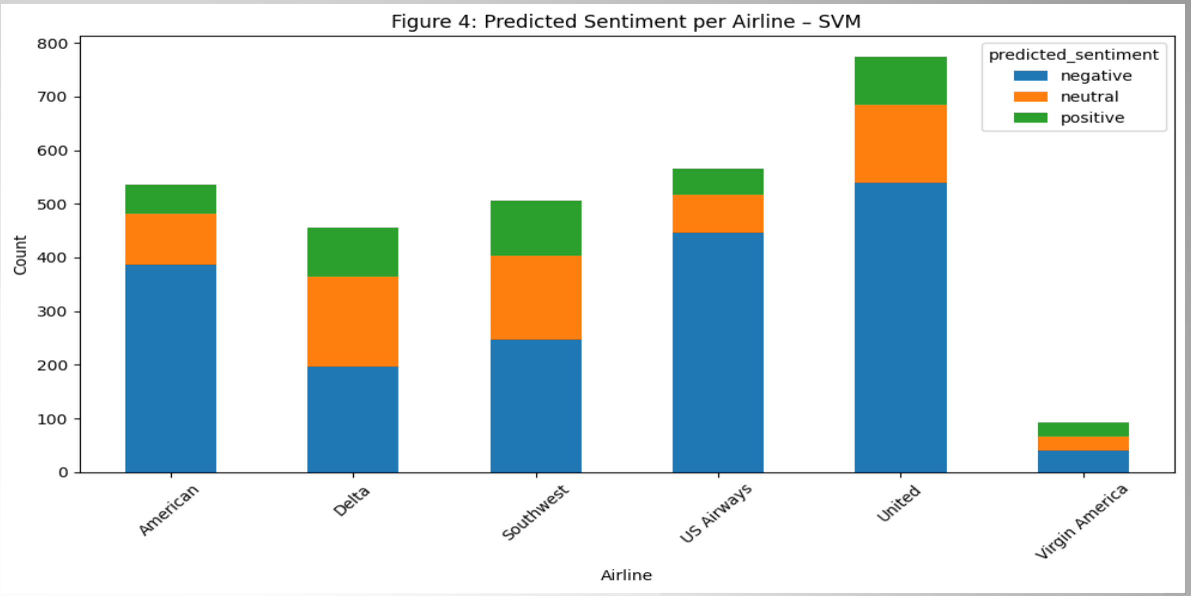
- Same test set for fair model comparison



RESULTS: CLASSICAL ML

SUPPORT VECTOR MACHINE (SVM)

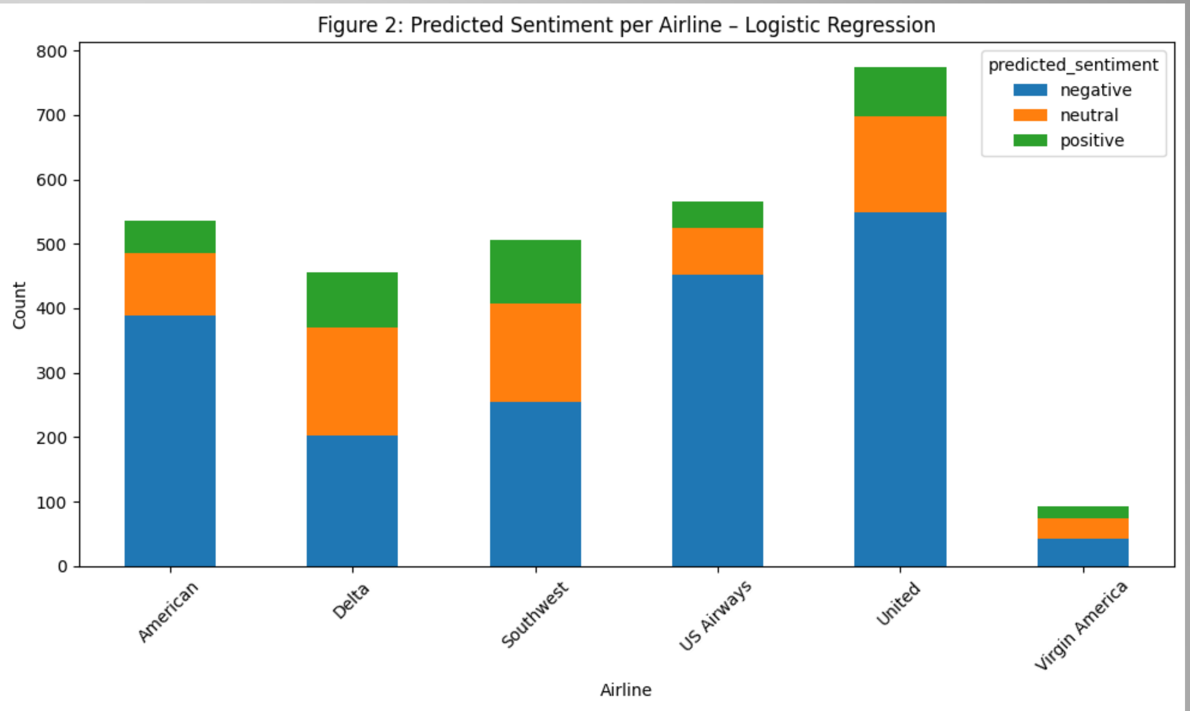
- Best performance: 93% accuracy
- Strong at differentiating negative vs. neutral sentiment




```
=== SVM Performance ===
Accuracy: 0.9279371584699454

              precision    recall  f1-score   support

    negative       0.99      1.00      0.99      1835
     neutral       0.82      0.87      0.84       620
    positive       0.84      0.72      0.77       473

    accuracy                           0.93      2928
   macro avg       0.88      0.86      0.87      2928
weighted avg       0.93      0.93      0.93      2928
```

LOGISTIC REGRESSION

- Accuracy: 92%
- Stable and consistent across classes




```
=== Logistic Regression Performance ===
Accuracy: 0.9183743169398907

              precision    recall  f1-score   support

    negative       0.97      1.00      0.99      1835
     neutral       0.80      0.87      0.83       620
    positive       0.86      0.67      0.75       473

    accuracy                           0.92      2928
   macro avg       0.88      0.85      0.86      2928
weighted avg       0.92      0.92      0.92      2928
```

NAIVE BAYES

- Accuracy: 79%
- Struggles with overlapping sentiment vocabulary

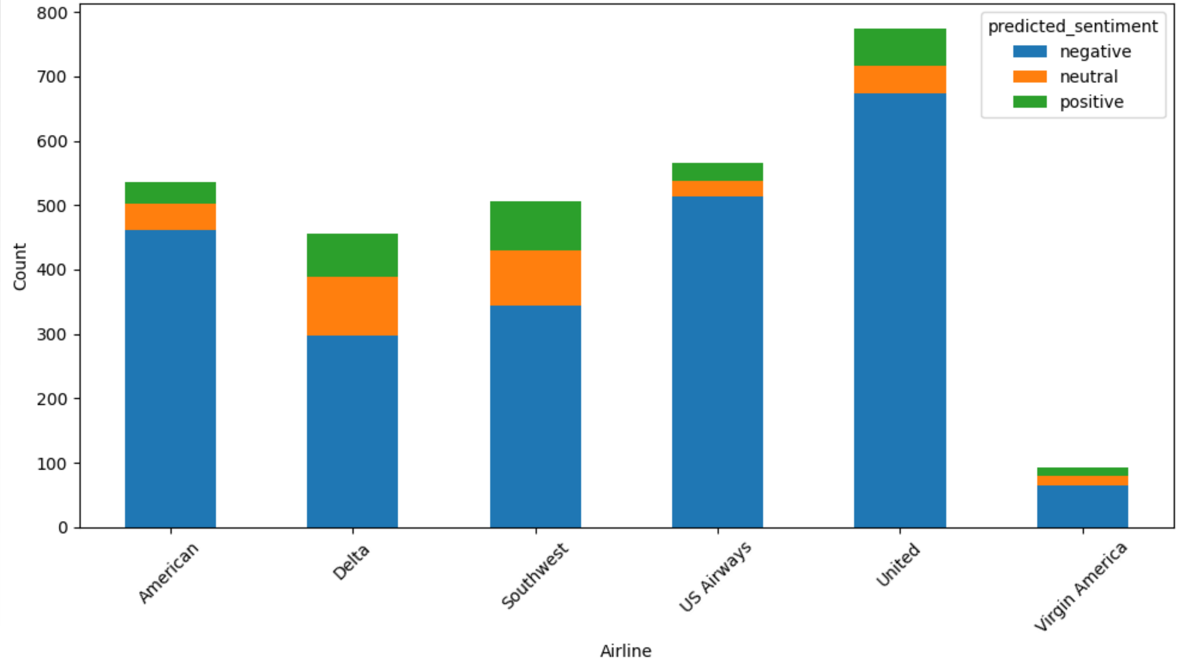
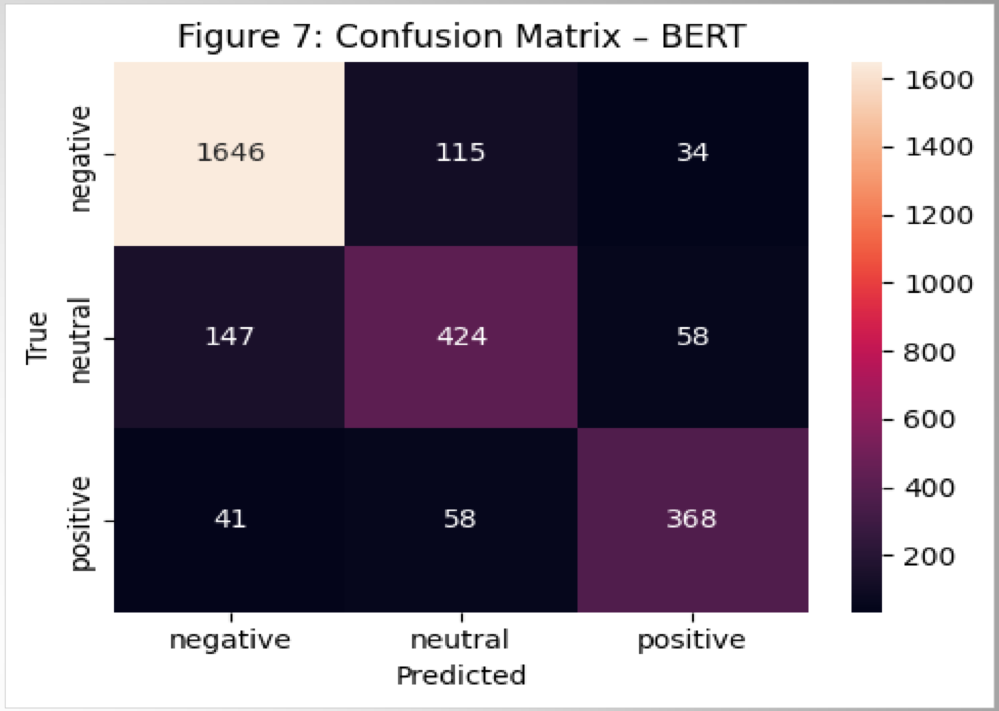


Figure 6: Predicted Sentiment per Airline – Naive Bayes

```
=== Naive Bayes Performance ===
Accuracy: 0.7920081967213115

              precision    recall  f1-score   support

    negative       0.78      1.00      0.87      1835
     neutral       0.81      0.39      0.53       620
    positive       0.90      0.52      0.66       473

    accuracy                           0.79      2928
   macro avg       0.83      0.64      0.69      2928
weighted avg       0.80      0.79      0.77      2928
```

RESULTS: BERT

- Accuracy: 84%
- Better at capturing positive and nuanced sentiment
- Balanced confusion matrix
- Requires more computation + careful tuning




```
=== BERT Classification Report ===
              precision    recall  f1-score   support

    negative       0.90      0.92      0.91      1795
     neutral       0.71      0.67      0.69       629
    positive       0.80      0.79      0.79       467

    accuracy                           0.84      2891
   macro avg       0.80      0.79      0.80      2891
weighted avg       0.84      0.84      0.84      2891
```

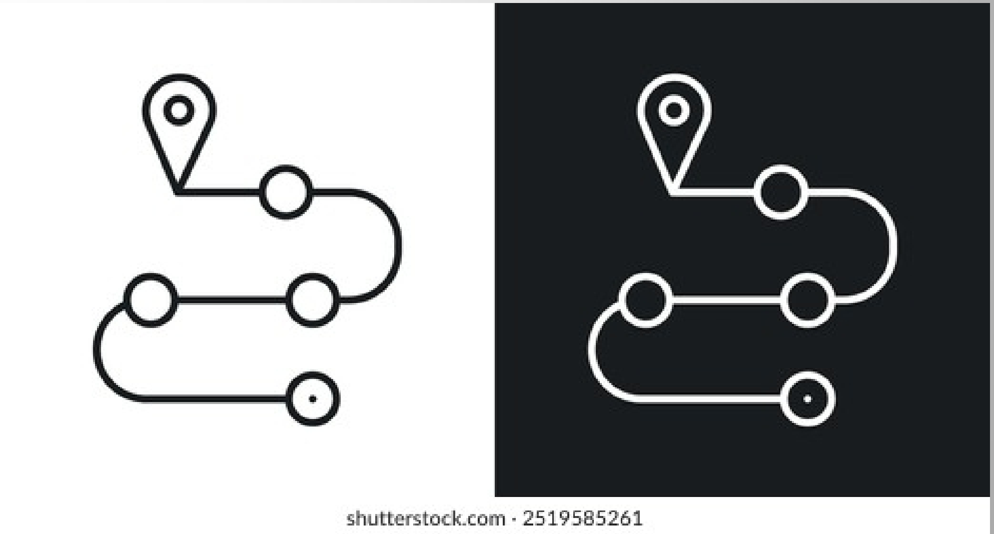
CONCLUSION & FUTURE WORK

Conclusion

- Classical ML models (especially SVM) perform extremely well with TF-IDF
- BERT provides deeper semantic understanding
- Best model depends on dataset size and complexity

Future Work

- Use larger and richer datasets
- Test more advanced transformers (RoBERTa, DistilBERT)
- Perform topic-specific sentiment analysis (delays, pricing, staff behavior)



REFERENCES

[1] A. S. Shitole and A. S. Vaidya, "Machine learning based airlines tweets sentiment classification," International Journal of Computer Applications, vol. 185, no. 20, pp. 32-35, Jul. 2023. [Online].

[2] F. Rustam et al, "Tweets Classification on the Base of Sentiments for US Airline Companies," Entropy (Basel, Switzerland), vol. 21, (11), pp. 1078, 2019.

[3] W. Aljedaani et al, "Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry," Knowledge-Based Systems, vol. 255, pp. 109780, 2022.

[4] M. D. Devika, C. Sunitha and A. Ganesh, "Sentiment Analysis: A Comparative Study on Different Approaches," Procedia Computer Science, vol. 87, pp. 44-49, 2016.

[5] G. Ravi Kumar, K. Venkata Sheshanna and G. Anjan Babu, "Sentiment analysis for airline tweets utilizing machine learning techniques," in International Conference on Mobile Computing and Sustainable Informatics, J. S. Raj, Ed. Switzerland: Springer International Publishing AG, 2020, pp. 791-799.

[6] L. Zhang and B. Liu, "Sentiment analysis and opinion mining," in *Encyclopedia of Machine Learning and Data Mining*, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, 2017, pp. 1152-1161.

[7] A. Hassan and A. Mahmood, "Deep learning for sentence classification," in 2017. DOI: 10.1109/LISAT.2017.8001979.

[8] J. Devlin *et al*, "BERT: Pre-training of deep bidirectional transformers for language understanding," Cornell University Library, arXiv.org, Ithaca, 2019.