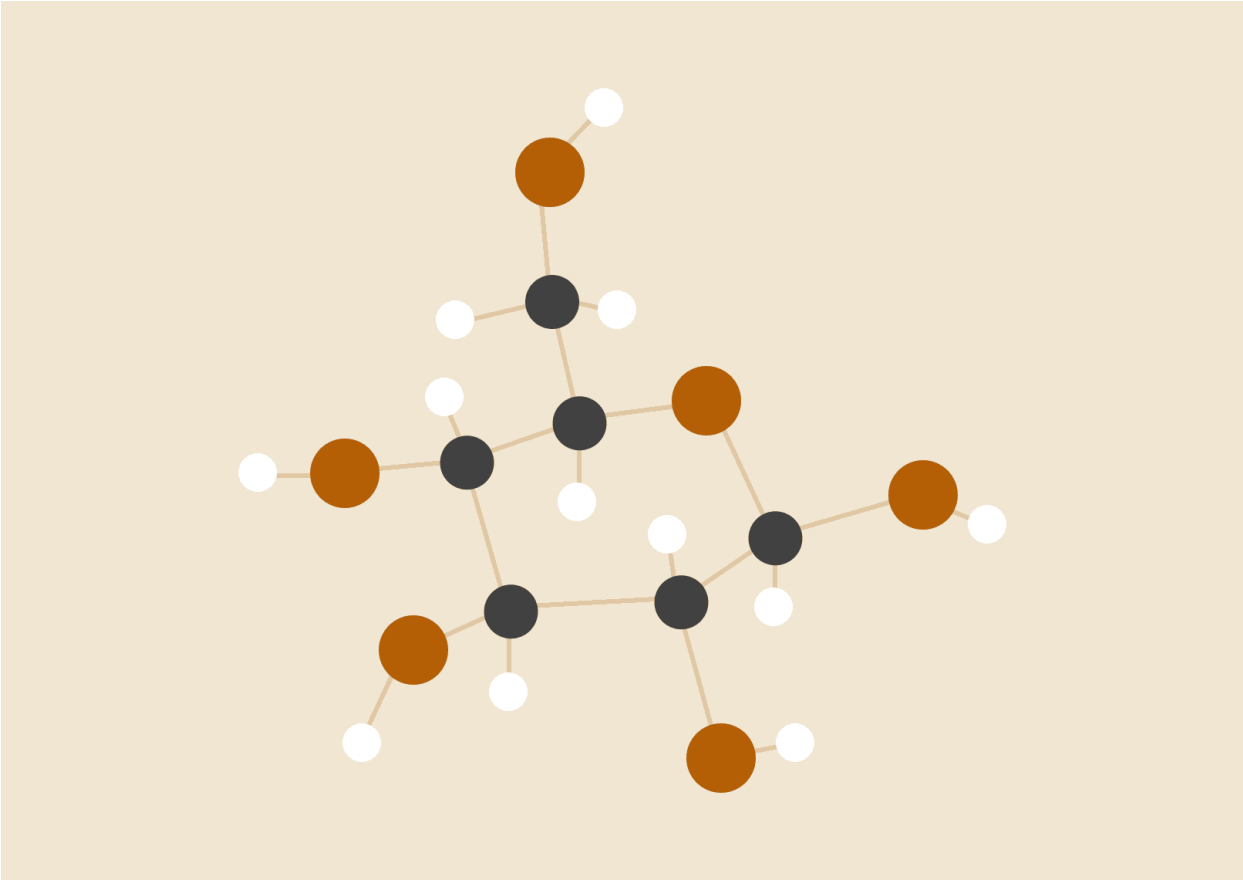
ML Mini-Project REPORT



Sarcasm Sleuth

OVERALL EVALUATION SUMMARY

MODEL / VARIANT	ACCURACY	PRECISION	RECALL	F1-SCORE	NOTES
RoBERTa (Default)	~53.6%	66.15%	50.59%	57.33%	Baseline transformer performance is lower.
RoBERTa + 3-Fold + Hyperparameter Finetuning	Fold Avg: ~59.7% Final: **65.2%**	72.8%	69.4%	Fold Avg: ~60.8% Final: **71.1%**	The final tuned model improved overall, especially F1.
RoBERTa + Extra Features & Context Embedding	~62.3%	~67.4%	~75.3%	~71.1%	Adding extra features helped boost recall and F1.
BERT (Scraped Data)	~60.9%	67.12%	62.03%	64.47%	Slightly better than baseline RoBERTa.
SVM (Linear Kernel)	~63.0%	80%	53%	~64.0%	Lightweight Model; competitive with transformers.
SVM (RBF Kernel)	~61.6%	86%	45%	~59.0%	Similar to linear SVM, with a bit lower F1.
XLNet	~60.9%	61.19%	59.42%	60.29%	Shows balanced but modest performance.

Key Takeaways / Insights:

- **Best Performer:** The **RoBERTa model with 3-Fold hyperparameter tuning (Final model)** yielded the highest accuracy (65.2%) and best F1 score (71.1%), suggesting that fine-tuning and using cross-validation can significantly boost performance.
- **Extra Features Impact:** Incorporating extra features and context embeddings into RoBERTa improved recall and F1, highlighting that additional context is beneficial for detecting nuanced sarcasm.
- **Traditional vs. Transformer:** While SVM models (both linear and RBF) achieve competitive accuracies (around 61–63%), they generally lag behind transformers in capturing the subtleties of language, which is critical for sarcasm detection.
- **Transformer Variants:** BERT and XLNet both provide moderate performance, with XLNet offering balanced metrics but not exceeding the tuned RoBERTa variants.