Beyond the Numbers: Leveraging MD&A Sentiment to Improve Financial Forecasting Accuracy

OVERVIEW

Our project initially began in partnership with the US Air Force Academy and the US Pacific Air Forces (PACAF) to explore how a fine-tuned large language model (LLM) could extract and summarize mission-critical information from unstructured military documents, such as site surveys and logistics reports, to support PACAF planners and decision makers.

However, compiling many unclassified documents resulted in creating a classified dataset ultimately limiting our ability to proceed with our project. Rather than abandon the project, we pivoted it to a new domain with similar challenges: financial analysis using publicly available Securities and Exchange Commission (SEC) filings.

These documents contain structured and unstructured data, require nuanced interpretation, and support high-stakes decision-making that make them an ideal analog to PACAF's operational documents.

Introduction

An LLM's ability to perform natural language processing (NLP) is transforming how people analyze unstructured data, especially documents. Financial analysts build forecasting models using data from SEC quarterly (10-Q) and annual (10-K) reports, but these text-heavy reports can be long and complex. LLMs can help enhance analysis through NLP techniques, like sentiment analysis.

All 10-Q/K filings contain "Management Discussion & Analysis" (MD&A) sections that provide commentary on financial results, risks, and future outlook. Sentiment analysis of MD&A sections can score the qualitative tone and generate quantitative metrics, such as positive/negative sentiment score.

Using a regression model, XGBoost, we forecast a company's future quarterly revenue based on past quarterly performance. We then compare our forecast's accuracy both with and without an MD&A sentiment score to test whether language-based metrics have a discernable impact on a model's prediction.

Methods

Data Collection – Collected 10-Q/Ks from the SEC's Electronic Data Gathering, Analysis, and Retrieval system.

Text Processing – Used Gemini 2.5 to extract financial data, such as assets, liabilities, net income loss, and revenue, from 10-Q/Ks and output this data into a CSV file.

Sentiment Analysis – Used Gemini 2.5 on 10-Q/K MD&A sections to develop a sentiment score ranging from -3 (Severely Negative) to +3 (Strongly Positive) to quantify executive's commentary tone.

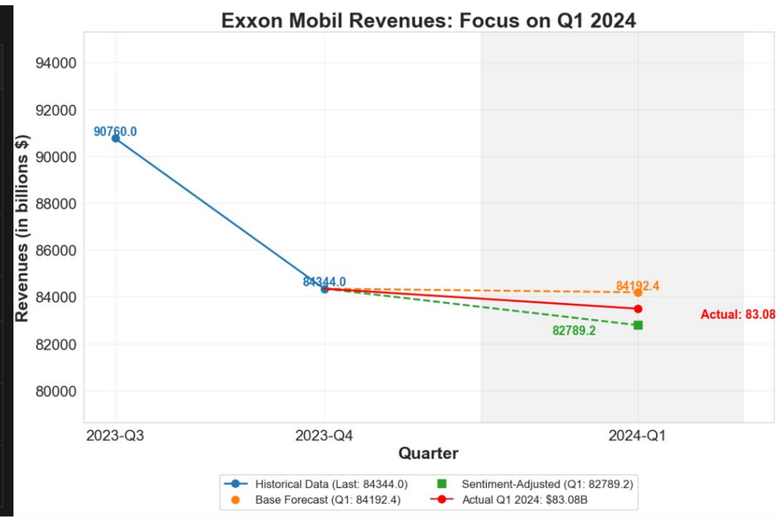
Predictive Modeling – XGBoost is well-suited for forecasting with small datasets of both structured numerical data (e.g., revenues) and engineered sentiment scores from unstructured text making.

Results & Conclusions

On the left is interpretable sentiment scoring table that maps our custom -3 to +3 scale to its label, example working, and coded language typically found in MD&A sections. On the right is an example of the effects of adding sentiment to our model. Adding our sentiment score allowed for a more accurate future quarterly revenue forecast.

Realistic Financial Sentiment Scale with Example Phrases

Score	Sentiment Label	Example Company Wording (Realistic)	Typical Coded Language/Keywords
-3	Severely Negative	"We are evaluating all strategic alternatives to address ongoing challenges."	"strategic alternatives," "material uncertainty," "challenging environment," "liquidity constraints," "substantial doubt"
-2	Negative	"The company continues to face headwinds impacting revenue and profitability."	"headwinds," "pressures," "challenging market," "declines," "adverse impact," "cost reduction initiatives"
-1	Slightly Negative	"Results were below expectations due to softer demand in key markets."	"below expectations," "softer demand," "margin compression," "modest decline," "temporary setback"
0	Neutral	"Performance was broadly in line with the prior period."	"in line with prior period," "stable," "no material change," "mixed results," "remained consistent"
+1	Slightly Positive	"We saw incremental improvement in operational efficiency this quarter."	"incremental improvement," "slight increase," "early signs of recovery," "modest growth," "encouraging trends"
+2	Positive	"The company delivered strong top-line growth and improved margins."	"strong growth," "improved margins," "outperformed expectations," "robust demand," "solid performance"
+3	Strongly Positive	"We achieved record revenues and are well-positioned for continued expansion."	"record revenues," "well-positioned," "exceptional results," "industry-leading," "significant momentum"



Born from the original PACAF project's goal of summarizing technical logistical reports to support decision making, we've shown how sentiment analysis can function as a proxy for executive tone, uncertainty, or optimism, ultimately helping to refine model accuracy and interpretability.

CONTRIBUTORS: Rama Deshpande, Harvey Trevino (UCCS), Ananya Manchiganti, Susan Dang (MSUD), Eric Almog, Esha Setty, Kayla Greathouse (MSUD), Leo Person. **TA**: Sriram Kiron

Acknowledgements: We thank the University of Colorado Colorado Springs (UCCS) and the Metropolitan State University of Denver (MSUD) for allowing their students to participate. We also thank Chief Master Sergeant William Baez for his guidance, and the USAFA cadets, Captain Guadalupe Hernandez, and the PACAF team for their support.

Data Mine of the Rockies
Spring 2025