

To: ShopNow Management Team

From: Marketing Analytics Department

Date: February 3, 2023

Subject: Improving Purchase Predictions through Customer Sentiment Analysis

Executive Summary

We want to see if customer feedback could help us make better predictions. Our usual approach was the Lasso model, which works with numerical data and gives us decent results. But we felt like we were missing something important. What is the actual sentiment and experience of the customer.

Adding sentiment analysis through VADER made a real difference. Our recall rate almost tripled, and the F1-Score went up as well. We weren't expecting such a big change. We also tried Hugging Face's sentiment tool, which seemed to pick up on things VADER missed. That gave us our best F1-Score at 40.86%.

What we learned is simple. The numbers matter, but so does what people are saying. When we combined both, our predictions got a lot better. Customer feedback isn't just useful for understanding satisfaction and it helps us build stronger models.

1. Baseline Sentiment Feature (VADER) Creation

To work with customer comments, we needed to convert words into numbers our models could utilize. VADER is a simple tool that reads through feedback and scores it based on the language people use every day. If someone writes something negative, VADER assigns a score closer to -1. Positive comments get scores closer to +1. Neutral feedback would be somewhere in the middle around 0.

Vader already knows how to interpret thousands of common words and phrases, so we didn't have to teach it from scratch. Once we had these sentiment scores, we could add them to our dataset just like any other variable. What started as subjective opinions became actual data points we could feed into our predictive models.

2. Predictive Modeling: VADER vs. The Original Baseline

When we added these VADER scores to our original Lasso model, the improvement was immediate and significant.

Metric	Original Baseline (Lasso)	VADER Integrated Model	Improvement (Lift)
Accuracy	77.61%	78.72%	+1.11%
Precision	51.58%	54.98%	+3.40%
Recall	12.07%	31.28%	+19.21%
F1-Score	19.56%	39.87%	+20.31%

Why this matters:

The baseline model we started with wasn't finding many of the customers we cared about—only around 12% of them. When we added VADER sentiment scores, that number shot up to 31%. That's a significant improvement for something as simple as including what customers were saying.

The F1-Score also went up quite a bit, more than doubling from where we started. Basically, this score tells us how well the model balances finding the right customers without flagging too many wrong ones. A higher score means we're getting closer to that balance, which makes the predictions more useful overall.

3. Exploring Alternatives: Hugging Face (HF) Comparison

While VADER is great, it mostly looks at words individually. We compared it to Hugging Face (HF), this tool reads whole sentences to understand context.

Metric	VADER-Based Model	Hugging Face Model	Difference
Accuracy	78.72%	78.61%	-0.11%
Precision	54.98%	54.29%	-0.69%
Recall	31.28%	32.76%	+1.48%

Metric	VADER-Based Model	Hugging Face Model	Difference
F1-Score	39.87%	40.86%	+0.99%

The Hugging Face model gave us our best overall results. It was slightly better at catching the subtle details that VADER sometimes misses.

4. Why the Tools Disagree: A Deep Dive

The reason for the difference is simple: Context. VADER counts words and Hugging Face understands sentences.

Customer Feedback	VADER Score	HF Score	Why the Difference?
"I am thinking of canceling... it doesn't seem worthwhile."	+0.34 (Positive)	-0.99 (Negative)	VADER saw "worthwhile" as a plus; HF understood the context of "canceling."
"The product is too expensive... doesn't justify cost."	0.00 (Neutral)	-0.99 (Negative)	VADER didn't find "angry" words; HF understood the dissatisfaction with price.
"Shipping was quick... exceeded my expectations."	0.00 (Neutral)	+0.95 (Positive)	VADER missed the positive nuance; HF recognized the successful outcome.

Final Conclusion & Recommendations

The big difference now is that our model doesn't just track customer's behaviors, but it also detects the sentiment of the customer on our product and service. That context matters more than we initially thought. When we looked at actual customer comments, we found patterns that our previous model would have completely missed. For example, when someone writes "I am thinking of canceling... it doesn't seem worthwhile," VADER scored it as positive because it picked up on the word "worthwhile." Hugging Face, on the other hand, understood the broader context and correctly flagged it as negative. The same thing happened with price complaints. The comments of "The product is too expensive... doesn't justify cost" came across as neutral to VADER, but Hugging Face recognized the underlying dissatisfaction. Even positive feedback got lost sometimes. A comment saying "Shipping was quick... exceeded my expectations" didn't register with VADER, but Hugging Face caught the satisfaction.

These differences matter because they help us understand which customers are at risk and which ones are happy, even when the language isn't obvious. The 20% improvement in F1-Score tells us that sentiment analysis isn't just a nice addition—it's something we should be using consistently. Going forward, our recommendation is to stick with the Hugging Face model. It's clearly better at reading the context of a customer's review. And that will help us make better predictions about customer behavior.

One area we're particularly excited about is using this for early intervention. Because Hugging Face picks up on subtle warning signs—phrases like "thinking of canceling" or complaints about cost—we can start flagging at-risk customers before they churn. This means we can focus our retention efforts on the people who need it most, rather than spreading resources across everyone. It won't be perfect, but if we can catch even a portion of these customers early, it's worth the effort. We'll also be able to identify our most satisfied customers more accurately, which helps us understand what's working and who we should be targeting for promotions and upselling. This feedback isn't just telling us who's unhappy, it's also showing us where our strengths are too.