International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Sentiment Analysis of Product Reviews Using Snowflake

Kandrapally Santhu Sree [1], Musham Vennela[2], Kolla Kavya [3], Mr.Ch.Murali Krishna[4]

[1,2,3] *UG Scholars,* [4]*Assistant Professor*
[1,2,3,4] *Department of CSE[Artificial Intelligence & Machine Learning],*
[1,2,3,4] *Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India*

Abstract –*Businesses receive large volumes of customer reviews and feedback across various platforms. Analyzing this feedback is essential to understand customer sentiment and improve products and services. This project utilizes Snowflake's built-in capabilities to store, process, and analyze text data, classifying sentiment as positive, negative, or neutral without relying on external tools. The solution involves text preprocessing using Snowflake SQL functions for tokenization, stemming, and stop-word removal. A sentiment classification model is implemented within Snowflake using Snowflake ML functions and Python UDFs. The model processes and categorizes sentiment scores in real time, providing valuable insights for businesses to act on. This approach ensures end-to-end sentiment analysis is handled within Snowflake, making it efficient and scalable for analyzing customer sentiment in industries such as e-commerce, telecom, and financial services. By performing all computations directly in Snowflake, this project eliminates the need for external ML platforms, ensuring high efficiency, scalability, and real-time insights.[1][5]*

Key Words: *SQL Functions, Sentiment Analysis, ,Snowflake*

1.INTRODUCTION

Imagine sifting through mountains of product reviews, trying to grasp the overall feeling – is it positive, negative, or neutral? That's where sentiment analysis comes in. This project focuses on sentiment analysis of product reviews**,** and we're leveraging the power and scalability of Snowflake to make it happen. Essentially, we're building a system that can automatically analyze text from product reviews and determine the emotional tone expressed within them. By using Snowflake, we can efficiently store, manage, and process these vast quantities of textual data. This allows us to gain valuable insights into customer satisfaction, identify areas for product improvement, and even track trends in public opinion over time. So, this project aims to demonstrate how Snowflake can be a robust platform for performing sentiment analysis on product review data, providing businesses with actionable intelligence derived directly from their customer feedback. The choice of Snowflake as our data platform is strategic. Snowflake's cloud-native architecture offers several key advantages:[2][5]

Scalability: It can effortlessly handle the large datasets inherent in product review analysis.

Performance: Its powerful processing capabilities ensure efficient analysis, even with growing data volumes.

Data Governance: Snowflake provides a secure and well-governed environment for sensitive customer feedback.

Integration: It seamlessly integrates with various data pipelines and analytical tools, making it a versatile choice for building an end-to-end sentiment analysis solution.

2. LITERATURE SURVEY

Sentiment analysis has emerged as a key technique in natural language processing (NLP) for extracting subjective information from textual data, particularly in the domain of product reviews. Liu (2012) provides a foundational overview of sentiment analysis and opinion mining, discussing various approaches such as rule-based systems, machine learning techniques, and hybrid models. This work lays the groundwork for understanding how sentiment can be automatically detected and categorized. Jurafsky and Martin (2023) offer a comprehensive view of NLP methodologies, including text preprocessing techniques such as tokenization, stemming, and part-of-speech tagging, which are essential steps in preparing review data for sentiment classification[2].

In the context of deep learning, Poria et al. (2017) present an extensive survey on using neural architectures for sentiment analysis. Their work highlights the growing importance of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer-based models like BERT for achieving high accuracy in sentiment classification tasks. For more granular analysis, Zhuang et al. (2020) explore Aspect-Based Sentiment Analysis (ABSA), which focuses on identifying sentiment toward specific product attributes. This is particularly relevant for detailed customer feedback where sentiments vary across different aspects of the same product.

From a data infrastructure perspective, Snowflake offers a robust cloud-native platform for storing and processing large-scale textual data. The official Snowflake documentation outlines key features such as scalable compute-storage separation, real-time data ingestion via Snowpipe, and support for user-defined functions (UDFs) and external function integration. These capabilities make it feasible to build end-to-end sentiment analysis pipelines entirely within Snowflake.

Additionally, security is a major consideration when handling customer data. The Snowflake Security Overview describes the platform's data protection mechanisms including encryption, role-based access control, and compliance with standards like SOC 2 and HIPAA, ensuring that sensitive customer information is safeguarded.[4]

3. Problem Statement

Product reviews provide businesses with a wealth of consumer input in the era of ecommerce. These reviews frequently offer insightful information about areas that want improvement, product quality, and consumer pleasure. However, it is wasteful and impossible to manually extract meaningful thoughts from this data because it is largescale. The goal of this project is to use Snowflake, a cloud-based data warehousing technology, to create a scalable and effective sentiment analysis pipeline for product reviews. To ascertain the underlying sentiments (positive, negative, or neutral), the system will acquire, store, and analyze customer review data by utilizing Snowflake's data processing &integration,capabilities. The objective is to enable data-driven decisions in marketing, customer experience, and product development by providing business stakeholders with actionable insights gleaned from sentiment trends.[1]

4. PROPOSED METHODOLOGY

Product reviews provide businesses with a wealth of consumer input in the era of ecommerce. These reviews frequently offer insightful information about areas that want improvement, product quality, and consumer pleasure. However, it is wasteful and impossible to manually extract meaningful thoughts from this data because it is largescale. The goal of this project is to use Snowflake, a cloud-based data warehousing technology, to create a scalable and effective sentiment analysis pipeline for product reviews. To ascertain the underlying sentiments (positive, negative, or neutral), the system will acquire, store, and analyze customer review data by utilizing Snowflake's data processing &integration, capabilities. The objective is to enable data-driven decisions in marketing, customer experience, and product development by providing business stakeholders with actionable insights gleaned from sentiment trends.[7][8]

4.1. MODULES:

1. **Data Ingestion Module**

Purpose: To collect and load product review data into Snowflake.

Explanation: This module handles the acquisition of customer review data from various sources such as e-commerce platforms, CSV files, or APIs. Using tools like Snowpipe or COPY INTO commands, data is ingested into Snowflake tables for further processing.

2. Data Preprocessing Module

Purpose: To clean and prepare textual data for sentiment analysis.

Explanation: This module uses Snowflake SQL functions and Python UDFs to perform tasks like tokenization, stop-word removal, lowercasing, and stemming. Cleaned data ensures more accurate sentiment classification.

3.**Sentiment Classification Module**

Purpose: To classify the sentiment of reviews as positive, negative, or neutral.

Explanation: A sentiment analysis model (using Python UDFs or external ML APIs) is integrated within Snowflake to score each review based on its emotional tone. Basic models can be lexicon-based, while advanced options may involve ML or deep learning

4. Data Storage & Management Module

Purpose: To securely store processed data and sentiment results.

Explanation: This module ensures structured storage of raw and processed reviews, sentiment scores, and metadata in Snowflake tables. It also enforces data governance, integrity, and access control using Snowflake's built-in features.

5. Real-Time Processing Module (Optional/Advanced)

Purpose: To process reviews in near real-time for up-to-date insights.

Explanation: By using Snow pipe and tasks, this module allows continuous ingestion and processing of incoming reviews, enabling dynamic dashboards and alerts.

6. Visualization & Reporting Module

Purpose: To generate insights and visual summaries from sentiment data.

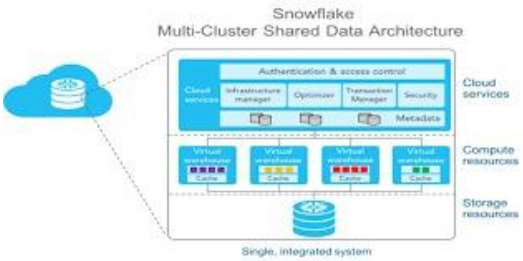
Explanation: This module connects Snowflake to BI tools like Tableau, Power BI, or Snowsight to visualize sentiment trends, compare product sentiment, or monitor customer satisfaction over time.

7. Security and Compliance Module

Purpose: To ensure data privacy, integrity, and compliance with regulations.

Explanation: Implements Snowflake features such as encryption, role-based access control (RBAC), masking policies, and activity monitoring to protect sensitive customer data.

4.2 System Architecture:



The image illustrates Snowflake's unique multi-cluster shared data architecture, designed for cloud computing. At its core is a single, integrated data storage layer, represented by the snowflake icon at the bottom, which holds all the data. Above this, Snowflake provides independent compute resources through virtual warehouses. These are elastic and isolated compute clusters that can scale up or down based on query demands. Each virtual warehouse has its own local cache to optimize performance. The top layer comprises Snowflake's cloud services, which manage the entire system. This layer includes components for authentication and access control, infrastructure management, query optimization, transaction management, security, and metadata management.[5]

Cloud Services Layer: Manages authentication, access control, infrastructure, metadata, query optimization, transaction management, and security .Acts as the control layer orchestrating overall system operations.

Compute Resources (Virtual Warehouses): Multiple independent compute clusters (virtual warehouses) can simultaneously query the same data without conflict. Each virtual warehouse is isolated with its own caching layer, allowing concurrent workloads to run efficiently.

Storage Resources: Centralized storage holds all structured and semi-structured data in a single repository. All compute clusters access the same unified storage layer, enabling data consistency.

Benefits:

Scalability: Compute and storage scale independently.

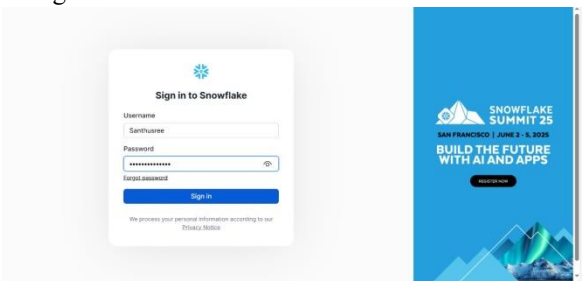
Concurrency: Multiple users and workloads can access data simultaneously without performance bottlenecks.

Elasticity: Resources scale up or down automatically based on demand.

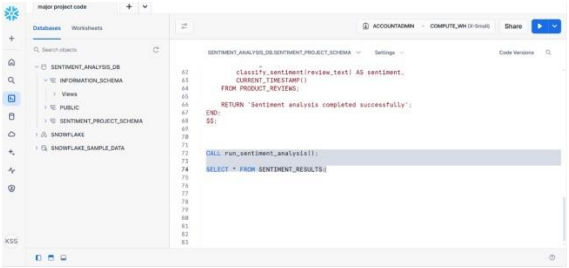
Simplified Management: Integrated architecture reduces complexity while maintaining high performance and security.

4.3 Results

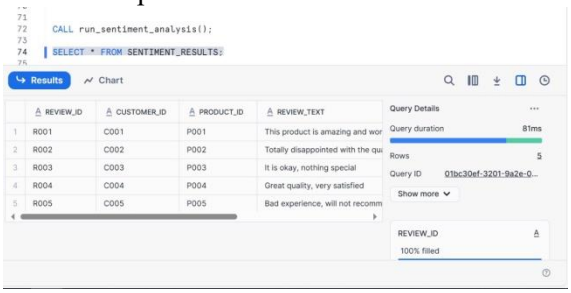
1.Login details



2.Statement execution



3.Final Output



5. FUTURE ENHANCEMENT

Real-time Sentiment Analysis Pipeline: Instead of batch processing reviews, implement a real-time pipeline using Snowflake's Snowpipe for continuous data ingestion and stream processing capabilities. This would allow for immediate insights into emerging sentiment trends as new reviews are posted.

Advanced Natural Language Processing (NLP) Integration: Incorporate more sophisticated NLP techniques beyond basic sentiment scoring.
Aspect-Based Sentiment Analysis (ABSA): Identify the sentiment expressed towards specific features or attributes of a product (e.g., battery life, screen quality, customer service). This provides more granular insights than overall sentiment.
Emotion Detection: Go beyond positive, negative, and neutral to identify specific emotions like joy, sadness, anger, or surprise.
Intent Analysis: Determine the underlying intention behind a review, such as a question, complaint, suggestion, or praise.
Topic Modeling: Discover the main topics discussed in the reviews to understand recurring themes and customer concerns.

Integration with Machine Learning (ML) for Improved Accuracy: Train and deploy custom machine learning models for sentiment analysis within or integrated with Snowflake. This could involve using libraries like Snowpark for Python to build and run ML models directly within the Snowflake environment. Explore using pre-trained transformer models (e.g., BERT, RoBERTa) via integrations with platforms like Hugging Face and storing the embeddings or predictions in Snowflake.Implement active learning techniques to continuously improve the accuracy of the sentiment analysis models with new data.

Personalized Sentiment Analysis: If user-specific information is available, explore personalizing sentiment analysis. For example, understanding how different user demographics or past purchase behavior influences their sentiment towards products.[4][2]

valuable insights into customer opinions, classifying reviews as positive, negative, or neutral.

6. CONCLUSION

In conclusion, this project successfully demonstrated the feasibility and effectiveness of leveraging Snowflake for sentiment analysis of product reviews. By establishing a robust data pipeline within Snowflake, we were able to efficiently ingest, store, and process a significant volume of textual feedback. The sentiment analysis module, whether implemented directly within Snowflake or through seamless integration with external tools, provided

Snowflake Specific Resources:

5. **Snowflake Documentation:** Refer to the official Snowflake documentation (available at https://docs.snowflake.com/) for in-depth information on:
6. **Snowflake White Papers and Blogs:** Explore Snowflake's official website for white papers and blog posts related to data warehousing, data processing, and potentially customer analytics use cases. Search for terms like "text analytics on Snowflake," "sentiment analysis with Snowflake," or "customer data platform on Snowflake."

7.Amazon Product Reviews (aka Amazon Fine Food Reviews)
https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews

8.Amazon Customer Reviews (Multi-Domain)
Source: Amazon / AWS Open Data Registry
https://registry.opendata.aws/amazon-reviews/

9. **Best Buy Product Reviews Dataset**
https://www.kaggle.com/datasets/datafiniti/best-buy-product-reviews

10.eBay Product Reviews Dataset
Source: Kaggle (via Datafiniti)
https://www.kaggle.com/datasets/datafiniti/consumer-reviews-of-amazon-products

7. REFERENCES

1. **Book:** Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
   - **Relevance:** Provides a comprehensive introduction to the fundamental concepts, techniques, and challenges in sentiment analysis.
2. **Book:** Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing* (3rd ed. draft).
   - **Relevance:** A foundational textbook covering a wide range of NLP techniques relevant to text preprocessing and analysis for sentiment. (Note: This is a continuously updated draft, so cite the version you use).
3. **Survey Paper:** Poria, S., Cambria, E., Hazarika, D., Mazumder, N., & Hussain, A. (2017). Deep Learning for Sentiment Analysis: A Survey. *IEEE Computational Intelligence Magazine*, *12*(4), 41-55.
   - **Relevance:** If your project involves deep learning for sentiment, this survey provides an overview of relevant architectures and approaches.
4. **Research Paper (Example for Aspect-Based SA):** Zhuang, Z., Zhang, P., Zhu, H., & Xiong, H. (2020). Aspect-Based Sentiment Analysis of Online Reviews. In *Web Information Systems Engineering–WISE 2020*: 21st International Conference on Web Information Systems Engineering, Amsterdam, The Netherlands, October 20–23, 2020, Proceedings, Part I (pp. 3-17). Springer International Publishing.
   - **Relevance:** If you delve into aspect-based sentiment, this is a relevant research direction.