# Delivering Insights from the Voice of Customer ML Models using **AWS EKS for a Leading Manufacturer**

Tiger Analytics partnered with the client to build a comprehensive model-driven solution, Voice of Customer analyzes the reviews by converting the reviews into embeddings using transformer models and then clustering them. This in-house solution was built to generate insights on the reviews submitted by the customers.

The Background

Our client is a US based leading manufacturer of industrial and home tools. Our client engaged us to

- Analyze customer data points, primarily e-commerce reviews, to generate valuable insights by leveraging transformer models and clustering techniques.

- Empowers end-users to analyze large volumes of customer reviews without manual efforts.
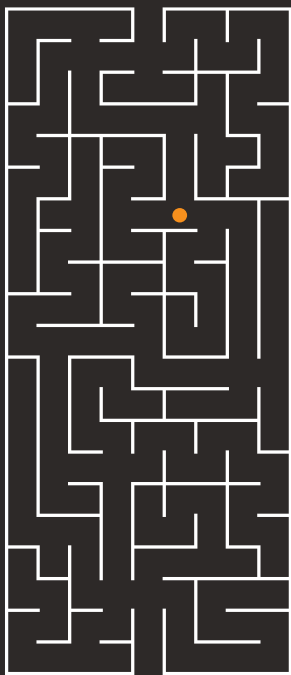
Key Challenges

/ The consumers are the Business Users who make informed decisions based on the results provided in the form of Power BI dashboards.

/ The consumers get a high-level insight without going through the large volumes of data manually.

/ The input data is collected from various e-commerce sites and undergoes a Brand Cleanup process.

/ Different variants of brand names are standardized and mapped to a consistent version for better analysis.

/ The client gets millions of reviews as input data. The client is concerned about: Perform the sentiment analysis around the product, What are the issues faced by the customers, What are the topics being discussed in the reviews..
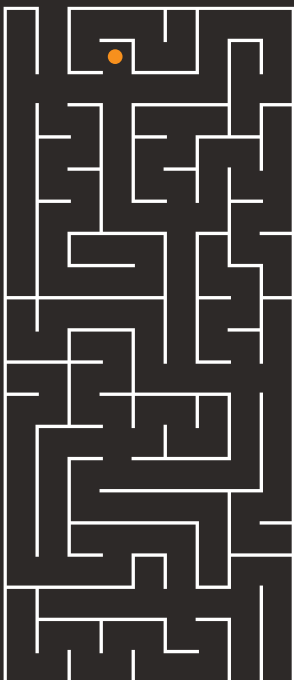
Our Solutions

Our Solution for the Voice of Customer analyzes the reviews by converting the reviews into embeddings using transformer models and then clustering them. Then the clusters go through an interpretation process where the highlights of each cluster are identified, and all this data with a bit of additional aggregation and processing is presented to the users on a Power BI dashboard.

This process is split into several steps and the logic is written as Python scripts. VOC uses several public language models to analyze the reviews, cluster them and to interpret the topics. The following diagram displays the VOC pipeline.

/ Brand Cleanup - As the data is scraped from multiple ecommerce sites, the data contains different variants for the same brand. We clean up most of these brands and map them to a standard version so that we can use that going forward.

/ Review Segmentation - The objective here is to split reviews into chunks if we find that a review talks about multiple subjects in its texts. This is done using SPACY to parse for POS tags and then look for inconsistency in the subjects mentioned across consecutive sentences. This is run on all reviews, split them into chunks and write them back to intermediate tables.

/ Embeddings Generation - We use a transformer model to generate embeddings for each review. We use UMAP algorithm to reduce the dimension of embeddings (which is 500+ for most algorithms) to a manageable number (around 10 dimensions)

/ Brand Cleanup, Master Data Prep, Category Cleanup (PySpark on Kubernetes)
/ Review Segmentation (PySpark on Kubernetes with Spacy English)
/ Embeddings Generation (Python on Kubernetes GPU with a fine-tuned version of XLM Roberta)
/ Review Clustering and Topic Interpretation (PySpark on Kubernetes)
/ Setup: The VOC pipeline runs as scheduled batch jobs with the following setup:
/ Code Versioning: GitHub for code management and collaboration.
/ Development Environment: JupyterLab for interactive data exploration and model development.
/ Orchestration & Scheduling: Jenkins to orchestrate and schedule the pipeline tasks.
/ Platform: AWS EKS (Elastic Kubernetes Service) for scalable and containerized deployment.

Tech Stack

/ AWS Cloud Platform: EKS Cluster

Solution Architecture
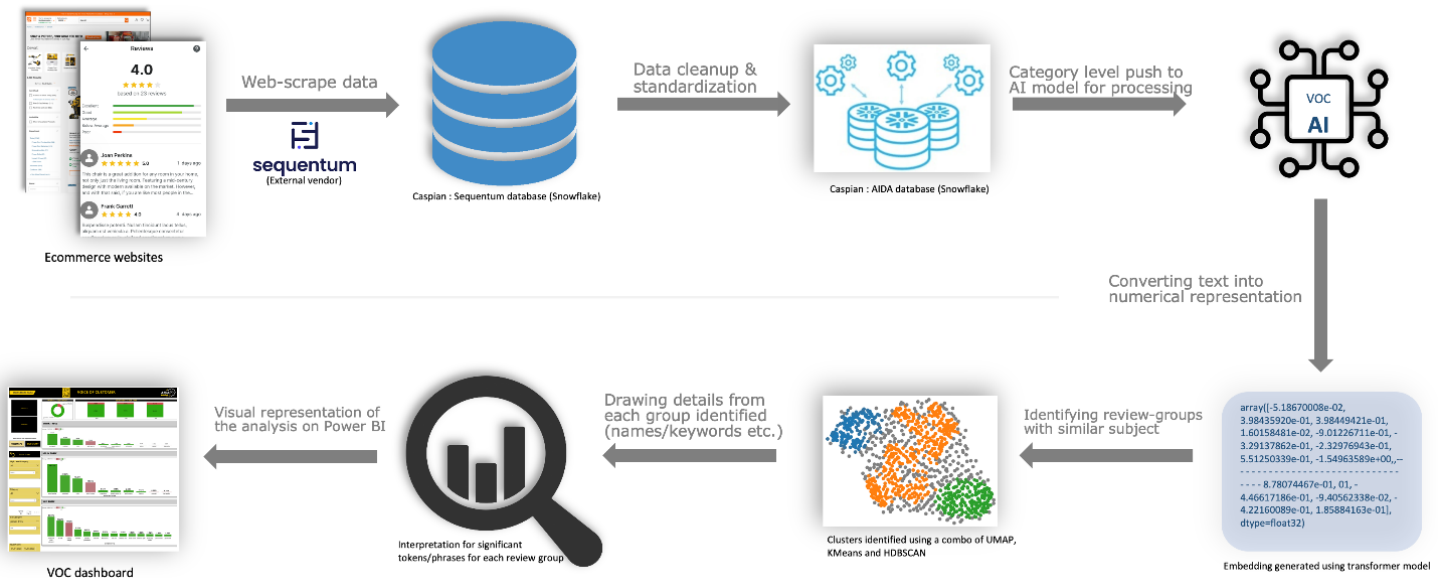
Value Delivered

▪ Accelerated insights on the various customer reviews/ aspects.
▪ Faster and efficient process
▪ The integration with Visualization tool allows for easy access and visualization of high-level insights, empowering business users to make data-driven decisions based on customer feedback.

About Tiger Analytics

Tiger Analytics is a global leader in AI and analytics, helping Fortune 1000 companies solve their toughest challenges. We offer full-stack AI and analytics services & solutions to help businesses achieve real outcomes and value at scale. We are on a mission to push the boundaries of what AI and analytics can do to help enterprises navigate uncertainty and move forward decisively. Our purpose is to **provide certainty to shape a better tomorrow.**

Being a recipient of multiple industry awards and recognitions, we have 4000+ technologists and consultants, working from multiple cities in 5 continents.

www.tigeranalytics.com