Tiger Analytics

Exploring Chat Intelligence

How NLP And Generative AI are helping businesses derive strategic insights from chat conversations

January 2024

Table of Contents

1. Introduction 3
2. Delivering Enhanced Business Value through Chat Intelligence 4
3. When Chat meets Gen AI and NLP 6
4. From Insight to Interaction: Tiger Analytics' Chat Intelligence 9
5. Overcoming Chat Mining Challenges 13
6. Transitioning to Deep Learning-Based Modelling and Fine-Tuning 16
7. Ablation Analysis 21
8. Transforming Conversations, Empowering Businesses 23
9. About the Authors 24

Introduction

Human interactions, whether over the phone or through chat, convey a wealth of information and signals. Harnessing these signals could lead to valuable insights that can power critical business decisions.

This whitepaper discusses the potential of leveraging Natural Language Processing (NLP) and Generative AI (Gen AI) techniques to derive valuable business insights from such interactions (chats) through Tiger Analytics' Chat Intelligence, the importance of mining chat data, the challenges that come with it, the applicability of deep learning models, and the relevance of incorporating industry and domain-specific nuances in such solutions.

It also explores how Tiger Analytics helped a major financial brokerage firm uncover hidden signals and cues in client conversations (chat) by deploying an NLP and Gen AI-based solution, resulting in enhanced business and operational visibility along with improved efficiencies.

Delivering Enhanced

Business Value

Through Chat Intelligence

A traditionally manual process, lead generation demands inventive approaches to keep pace with digital transformation. Innovative platforms, such as Live Chat, allow organizations to interact directly with customers who visit their digital platforms. Customers can interact with a chat box in real-time to access the information they need about the company, product(s), or service(s).

Whenever a customer interacts with the Call Center, the individual gets connected with a specialized executive, who discusses and solves the queries of that individual. The notes from that chat conversation are then carefully monitored for quality improvement purposes, as well as to generate any relevant insights. At this point, organizations could face the following challenges:

Not all chats get monitored and analyzed within a specific period - For example: **A Fortune 500 financial brokerage company takes nearly 150 mins to manually review 1% of its daily chats.**

When they do get monitored, in organizations where the chats are passed onto a dedicated team, it takes a long time to analyze 100% of the chats. For example: **A leading MNC takes 15,000 mins to manually review the chat conversations that are logged in a single day.**

The physical process of analyzing the chats often overlooks semantics and fails to capture the vital signs. For example, a chat conversation that reads **"I have $3 million in my Individual Retirement Account (IRA) and I am facing difficulties managing my portfolio" is often overlooked in the process of solving an operational problem.**

To address the challenges and help organizations, we require a solution that has the ability to answer a critical question:

How can we efficiently monitor all the chat conversations and unearth vital signs relevant to the business growth?"

"Site visitors who use web chat are 2.8x more likely to convert than those that don't.

FORRESTER

"A buyer who chats will spend 60% more."

BOLD 360

The constant market push to boost revenue emphasizes the importance of incremental sales. Taking the above statistics into consideration, it is crucial for organizations to invest in and build a chat engine that not only interacts with customers, but also gathers insights about their behavior

When Chat meets Gen AI and NLP

CHAT INTELLIGENCE refers to our specialized technology and solutions that leverage NLP and Gen AI capabilities. At its core, Chat Intelligence encompasses the use of advanced AI-driven algorithms to enhance chat and messaging systems. **These systems can understand, interpret, and generate human-like text, based on natural language input, resulting in more sophisticated and valuable user interactions**

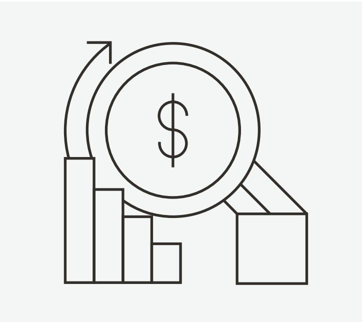
In the realm of digital communication, the use of sophisticated, context-aware models forms the core of the Chat Intelligence ecosystem. It leverages the capabilities of Large Language Models (LLMs) and their ability to analyze patterns and generate text to offer an enhanced experience for the user. Chat Intelligence enables users to converse in a human-friendly language, offering real-time intent identification and conflict resolution. It also generates valuable business insights that can help highlight inefficiencies in existing processes

Furthermore, over time, Chat Intelligence can identify patterns in the data to help isolate Frequently Asked Questions (FAQs), detect complaints and frauds, and understand user feedback.

It can help in offering new products and services for the user to interact with that they may not have used previously. This not only provides a unique and personalized experience for the end user but also streamlines and adds value to the existing business processes.

Driving value across industries with

Chat Intelligence



FINANCE

i. Identify opportunities or leads for incremental business and cross-selling

ii. Identify areas of customers’ concern and improve Customer Experience (CX) score

iii. Identify customer purchase intent and trigger relevant marketing intervention strategies

iv. Detect fraud

RETAIL

i. Identify opportunities for growth throughpersonalized product recommendations and suggestions based on user prompts

ii. Provide 360-degree brand feedback

iii. Develop ideas for product development.

INSURANCE

i. Identify signals of renewal opportunities

ii. Identify signals of retention opportunities for non-renewal customers

The AI Engine
behind Chat Intelligence systems

LLMs trained on a large corpus of text perform extremely well on general tasks. They show an unmatched ability to understand and analyze patterns and generate text. However, they may not perform well on the domain data, which is essential to have an effective enterprise-level chat solution. To overcome this, the models are retrained and fine-tuned, enabling them to learn the domain data and perform better on specific tasks.

By training models on such data, they develop the ability to understand domain and industry-specific terminologies and utterances. Incorporated into the Chat Intelligence ecosystem, these models help answer user queries in a better and more coherent way, leading to improvements in user satisfaction and operational efficiencies

Chat intelligence helps drive incremental business opportunities by identifying:

- New leads for businesses
- Signals from existing customers for additional business opportunities
- Potential customer dissonance triggering retention measures
- Themes for personalized marketing campaigns
- Upsell or Cross-sell opportunities
- Indicators or patterns that lead to fraud
- Customer retention strategies
- Customer sentiments

From Insight to Interaction
Tiger Analytics' Chat Intelligence

At Tiger Analytics, we help organizations provide a seamless and contextualized self-help experience to their users and generate insights into consumer behavior through our comprehensive NLP and Generative AI-driven Chat solution. We engage with our clients across various aspects of the solution development lifecycle, leveraging a holistic approach that includes strategy formulation, technical implementation, and ongoing assistance.

Our expertise encompasses a range of areas related to chatbot development, with a strong focus on virtual assistant creation, and the ability to identify and measure key metrics that define the success of chatbot implementations.

We harness cutting-edge technology and innovation to drive value for our clients. This involves leveraging open-source, enterprise as well as contextual LLMs such as Dialogflow, GPT-2, BloombergGPT, Alpaca, BARD, LLaMA, GPT-4, PaLM, Vicuna, BLOOM, and BERT and its different variants (like FinBERT, SciBERT, and ClinicalBERT). Every solution is carefully crafted and tailored to our clients' unique needs, ensuring delivery of maximum value and benefit from our solutions.

A walk through the chat journey

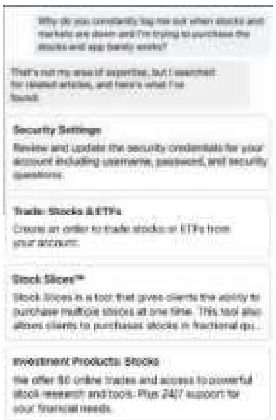
(A continuously evolving solution)

Chat Session Begins

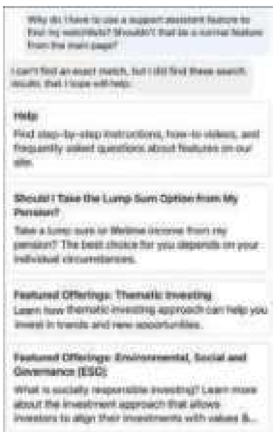
Conversational Agent

Program that answers frequently asked questions, troubleshoot issues and even make small talks

Identify the intent and display relevant options



Identify the difference between a query and a compliant



Deduct non-chatbot related feedback

Chat Session Ends

Post Chat Session

Chat Summarization

Using AI and NLP to understand and reduce a long text into a shorter summary while retaining the key information

Identify signals of incremental business

I am retiring at the end of this month. Can you please help me with the following?

Help in Customized Campaign

I am looking for information on a few things – Do you offer treasury bonds and bills?

Identify Cross-Sell Opp.

I'd like to buy another money market account that us not tied to the advisor

Identify Signals for Retention

Am thinking of moving my finds out to xyz

Identify Frauds Signals

Can you help me moving my fund to my foreign account?

Integrating chats with transactions

Business insights

NLP based conversational layer to analyze data and solve a use case. This helps reps gather client knowledge even before the client can speak to them

Here are a few areas, where we can help organizations unlock the potential of NLP and Gen AI-driven chatbots

Virtual Assistant Creation

Our team builds and designs virtual assistants that effectively cater to the needs of clients and end-users. We create chatbots that provide seamless and human-like interactions, enhancing user engagement.

Development Stages

From ideation to implementation, our team covers the entire spectrum of chatbot development. This includes defining the chatbot's purpose, functionality, and user interface design. Our expertise ensures that chatbots align with the client's objectives and target audience.

Performance Measurement

A critical aspect of any chatbot's success is its performance. We devise and implement effective metrics to measure chatbot performance. These involve evaluating user satisfaction, response times, and the bot's ability to understand and fulfill user requests. By analyzing these metrics, we can provide insights into how well the chatbot is meeting its goals.

Optimization and Iteration

Our team's commitment to ensuring successful chatbot implementation goes beyond the initial launch. We analyze user interactions and feedback to identify areas for improvement.

Through continuous optimization and iteration, we enhance the chatbot's capabilities and user experience over time.

Technological Proficiency

We leverage cutting-edge technologies and methodologies in our chatbot solutions. This includes NLP, Machine Learning (ML), and AI-driven techniques to create chatbots that are both intelligent and adaptable.

Customized Solutions

With our experience working with a wide variety of clients, our team at Tiger Analytics possesses a deep understanding of various industries and their unique challenges. This allows us to tailor chatbot solutions that align with specific industry requirements.

A walk through the chat journey
(a continuously evolving solution)

1

Development Stages

- Designing conversational flows
- Entity/intent models (conflict resolution, etc.)
- Measuring/improving NLU model performance

2

Enterprise Integration

- Augmenting chatbot with Enterprise search
- Featured answer widgets Results with fallback options
- Improving Live Agent experience

3

Functional Augmentation

- Improving customer engagement via personalization
- Complaints detection
- Customer feedback analytics
- Fraud detection

4

Enhancements

- User adoption through journey analytics
- Chat monetization
- Cross-sell - Upsell
- Retention

Overcoming

Chat Mining Challenges

NLP techniques have evolved rapidly in the last 5 years. New models are able to outperform traditional models on tasks such as classification, question-answering, and summarization. In this section, we will discuss the limitations of the traditional approaches and how at Tiger, we leverage the power of deep learning to overcome such limitations and build better solutions

Traditional approach

TF-IDF + Classifier

Deep Learning: Iteration-1

USE Embeddings + Classifier

Deep Learning: Iteration-2

Fine-tuned BERT Model

Traditional chat mining approaches often rely on TF-IDF and Word2Vec for feature extraction, followed by standard classification models.

However, these approaches suffer from challenges such as limited contextual understanding and high computational requirements.

Challenges
in the Existing Solution

I. Feature extraction process

TF-IDF is not very good at understanding context across an entire sentence or paragraph. Its memory footprint and time complexity increase as new data is incorporated.

II. Model training process

SVMs (especially for high dimensional data) and tree-based methods are the standard solutions developed. However, for textual data, deep learning-based models have proven the most effective. The SVM model cannot give out probabilities and tree-based methods require a lot of tuning with several moving parameters.

III. Model Maintenance

Quite frequently, a separate model is built for every individual source of data which makes the model maintenance very complicated. Separate models mean that there are too many metrics to track and it is difficult to dedicate resources to perform timely maintenance on each of them.

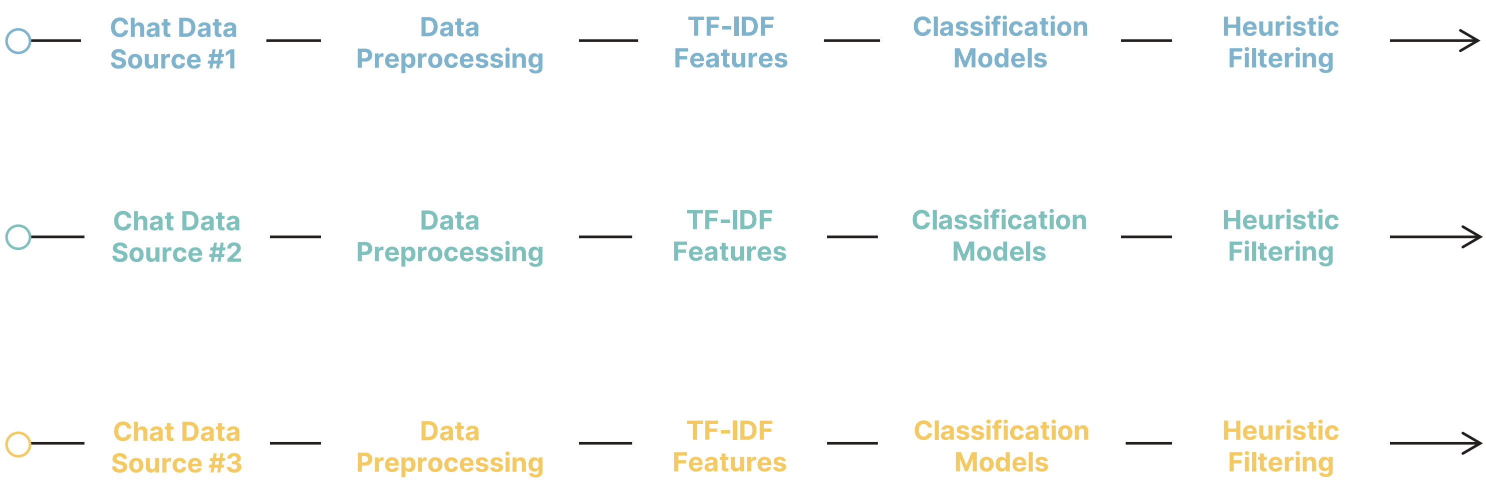
The Tiger Approach

A Progressive and Enhanced Solution

TF-IDF features are generated for different n-grams, where n is typically between 2 and 5. These features are followed by a standard classification model such as an SVM or random forest. Different thresholds are evaluated based on multiple factors, including model performance and business requirements.

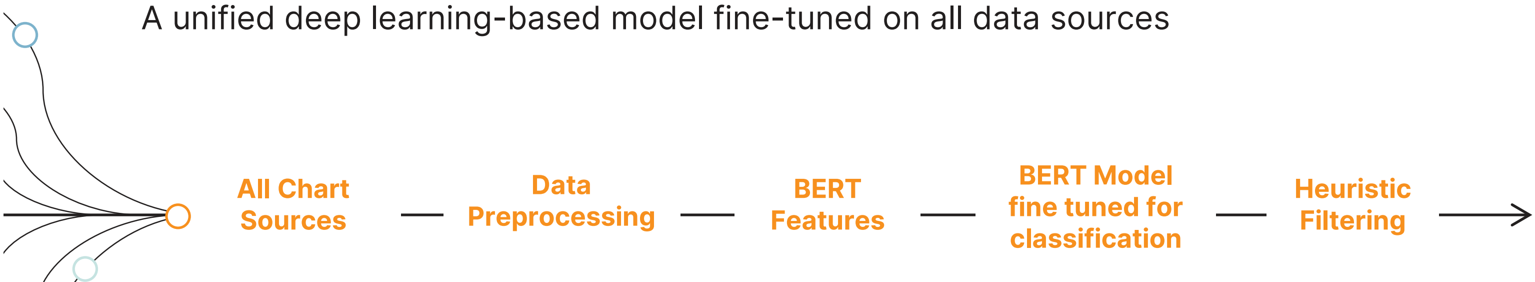
Commonly existing solution(s)

A separate model for every chat source



Tiger Analytics' solution architecture

A unified deep learning-based model fine-tuned on all data sources



This approach not only enhances model metrics through fine-tuning, but also improves various aspects of data engineering, encompassing maintenance, deployment, and more.

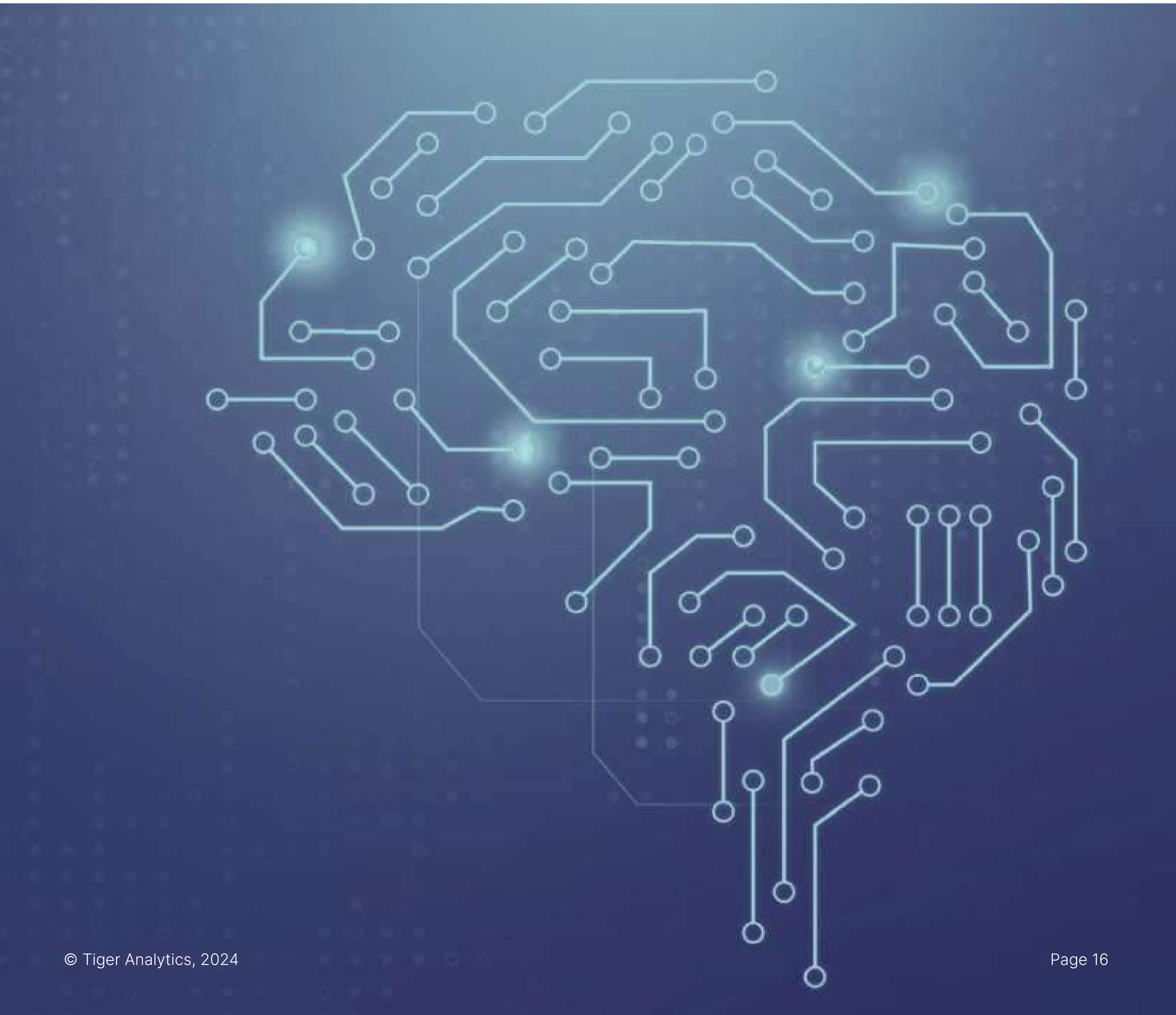
Transitioning to Deep Learning

Based Modeling and Fine-Tuning

Deep learning-based language models have made incredible progress and they can be used to overcome some of the shortcomings of the traditional methods. These models have millions of parameters that can be specifically tuned for the data and task at hand.

In the following section, we'll quickly explore the use of Universal Sentence Encoder (USE) Embeddings and the BERT model:

[Caveat: In all the methods, it's important to refresh the dataset and use the latest data for training.]



Iteration-1

Experimenting with USE Embeddings

USE embeddings is a pre-trained model developed by Google that converts text data into fixed-sized embeddings (usually 512-dimensional) to be used in various natural language understanding tasks. It can effectively encode semantics and sentence level information, making it useful for tasks such as sentiment analysis, text classification, and semantic similarity. These embeddings capture the semantic meaning and contextual information of the input text, allowing for various downstream natural language processing tasks.

At Tiger Analytics, we use

Transformer-based Architecture so that the USE model can generate embeddings that capture the complex semantic information of sentences, and

Deep Averaging Network (DAN) Architecture because it is computationally efficient and provides good performance in various NLP tasks.

Both USE architectures have been pre-trained on a large corpus of text data, enabling them to capture general semantic patterns. These pre-trained models can be perfected on specific tasks or used as feature extractors to transfer the learned knowledge to downstream applications such as sentiment analysis, text classification, or semantic similarity tasks. The output is a single 512-dimensional vector for any sentence

Advantages of using this approach include:

Improving

- Model accuracy and robustness
- Model acceptance rate
- Contextual features (USE embeddings can understand the context across the entire sentence instead of just its neighboring words.)

Reducing

- The training time of the model
(The dimensions of the features were fixed - Additional data added to the training did not affect feature dimensions.)

Limitations of using this approach include:

- USE is an improvement over the TF-IDF. However, the USE model was trained on a wide variety of data and not specifically for the domain data.
- It is not as robust as other neural network architectures and did not achieve the expected metrics from a business standpoint.
- The problem of using USE merely as a feature extraction tool was the majority of the weights of the network were frozen and only new parameters were learned. This resulted in a suboptimal model

Iteration-2

Experimenting with BERT Embeddings

Bidirectional Encoder Representations from Transformers (BERT) is a pre-trained language model developed by Google that has shown state-of-the-art results on various natural language understanding tasks. BERT is designed to capture both contextual and semantic information from text by using a bidirectional Transformers architecture.

It can be fine-tuned for specific tasks such as sentiment analysis, text classification, named entity recognition, and question-answering. The model learns to predict missing words within sentences, known as masked language modeling (MLM), and also learns to determine the relationship between sentences, known as next sentence prediction (NSP).

This pre-training allows the model to learn contextual representations that capture the nuances of language. Once pre-training is complete, BERT can be fine-tuned on specific downstream tasks with labeled data. By applying task-specific architectures and a relatively small amount of labeled data, BERT achieves state-of-the-art performance on a wide range of NLP tasks, including sentence classification, named entity recognition, and question-answering.

Fitting the model involves two stages

Mask Language Modeling

This is where the model learns to predict missing words within the sentences. This helps the generalized BERT model fit on your own dataset. For example - BERT is trained on a large dataset which is too generic for a specific use case. It is therefore better to retrain the masked language model on the domain-specific data. To perform this retraining, the original BERT model and its weights are initialized and a new MLM is fit using only the domain-specific data. In this training, all the BERT model's weights are updated. At the end of the training process, the new BERT model has the same architecture as the original BERT model but its weight has been updated for a specific domain. Now, this retrained model can be used for multiple downstream tasks.

Fine-Tuning

A classification model is fitted on top of the retrained model and this process of fitting the model is referred to as 'fine tuning'. This classification model is similar to the classification models discussed in the previous section. However, the beauty of this deep learning model is that all the weights of the BERT model are updated for this classification task thus making it very robust and accurate.

The outcomes of using this approach include:

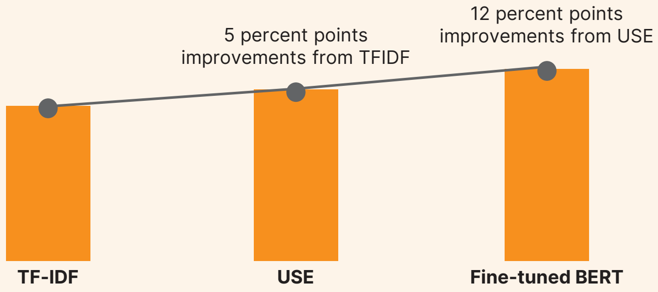
Improving

- Model accuracy and robustness
- Model's acceptance metrics by the business
- Volume of acceptance (TF-IDF algorithms were not able to capture all the opportunities)

Reducing

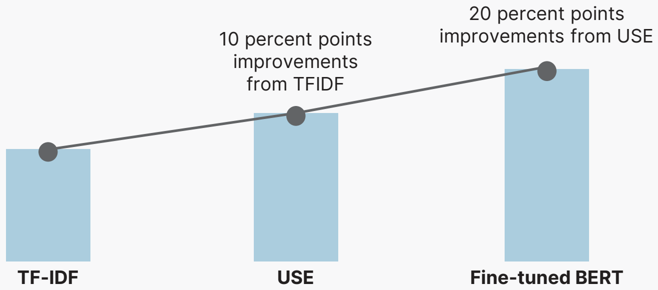
- False positive rates
- Unnecessary human effort in identifying the false positive cases

Model Metric Improvements



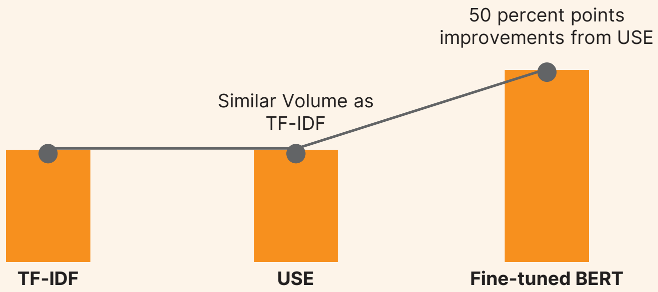
The accuracy of the model improved when the feature extractor moved from TF-IDF to a neural network. Further improvements were observed with the fine-tuning of all the parameters of the BERT model.

Business Metric Improvements



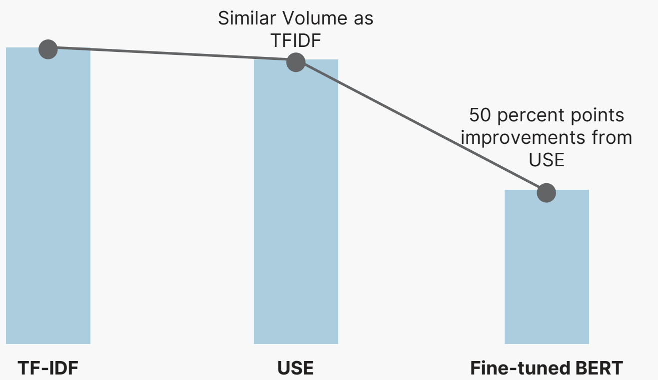
The business acceptance percentage measure of the model output acceptability by the end user increased dramatically by replacing the vanilla TF-IDF model with the fine-tuned BERT model.

Volume Generation Improvements

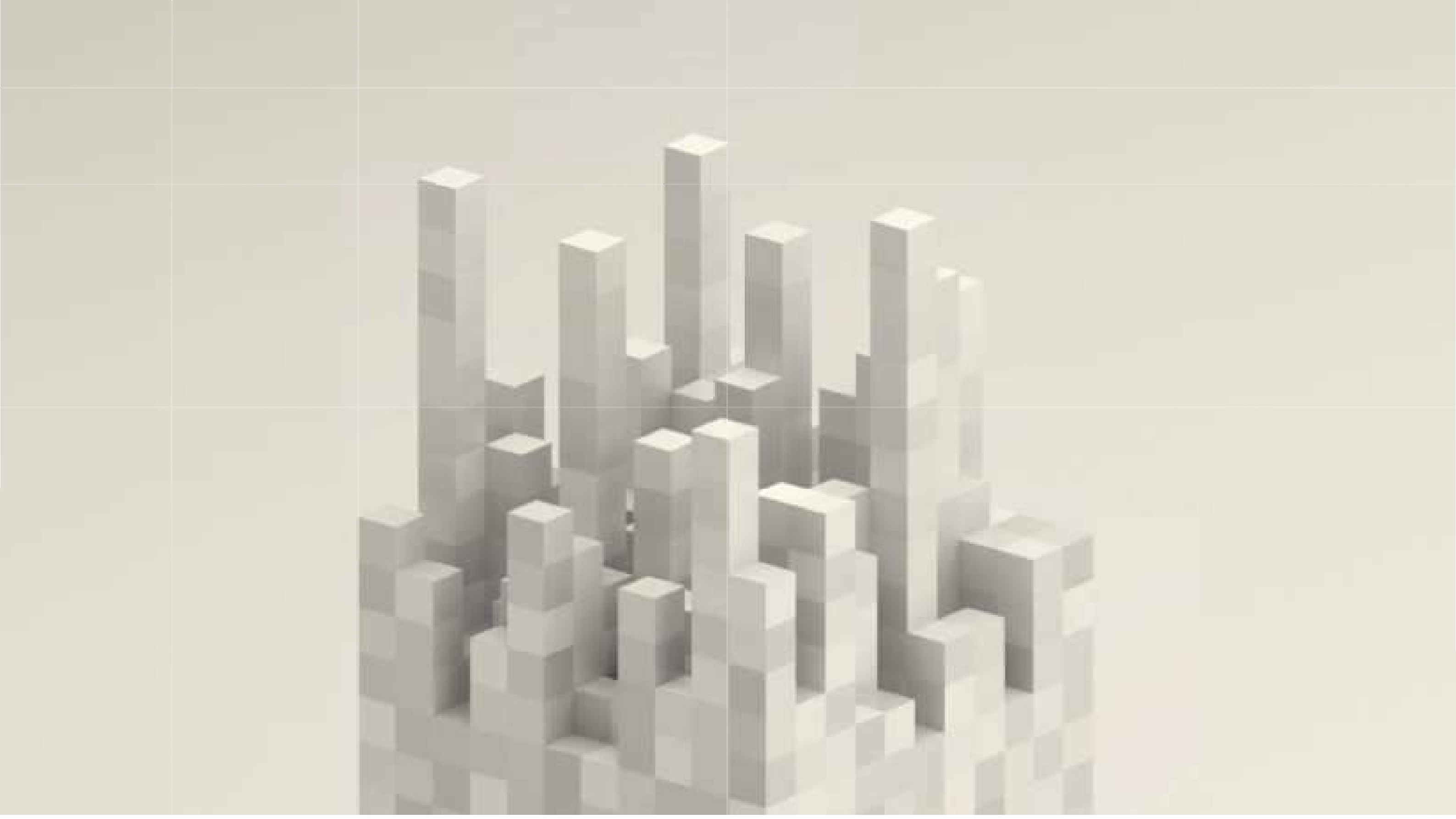


Beyond the accuracy and acceptance, the total chats crossing the model's threshold was the greatest in the fine-tuned BERT model, which shows that the TF-IDF model and the USE model were not able to capture all possible opportunities in the data and were not able to identify the entire potential.

Reduction in False Positive Rates



The false positive rate of the model was reduced. The TF-IDF and USE-based models were not able to fully differentiate between the business 'coming in' to the system and business 'going out' of the system. However, fine-tuning the BERT model helps capture nuances in utterances that are not possible by vanilla feature extractors.



Ablation Analysis

Ablation Analysis showcases the effectiveness of various models in discerning essential signals, illustrated through a series of examples. It highlights BERT's proficiency in capturing intricate meanings and contexts, resulting in the reduction of false positives and the enhancement of accurate identification of pertinent leads.

Text	Ground Truth	TD-IDF Prediction	USE Embedding Prediction	BERT	Explanation
Hi, I am considering moving all my accounts held at an outside firm to your firm.	Positive	Positive	Positive	Positive	A typical true positive that all models pick up
Hello, I am considering moving my account to a different firm	Negative	Positive	Positive	Negative	A typical false positive for the TF-IDF and USE embedding features
May I get some help. I am looking to open a new account and start contributing to it	Positive	Negative	Negative	Positive	Missed Opportunities that the BERT model picks up

1. In the first example **“Hi, I am considering moving all my accounts held at an outside firm to your firm.”**, the indication of the movement of money from external firms is clear and all the three models are able to pick up the signal of an incoming transfer.

2. In the second example, “**Hello, I am considering moving my account to a different firm.”**, the TF-IDF model and the USE embeddings-based model were not able to understand the nuances of the sentence. These are the typical false positives that the model struggled to differentiate:

In the case of TF-IDF, the n-grams formed would be of the type ‘moving my account’, ‘considering moving’, etc. These are typical signals from a positive class sample and the model mis-classifies it as a valid signal.	In the case of the USE embed-dings, the base model’s layers are frozen and only additional layers are trained. Since the base model was not developed on such data, the model is unable to understand the nuances of the utterances. The models pick up the ‘transfer’ signal but struggle to understand its ‘direction’	This is where the BERT model is able to outperform the other methods because of the domain knowledge it has from the retraining and fine tuning. The BERT model is able to distinguish between incoming and outgoing transactions.

3. In the third example, **“May I get some help. I am looking to open a new account and start contributing to it.”**, the TF-IDF and USE model’s output probabilities are below the threshold and hence are lost opportunities. However, the BERT model’s fine-tuning helps rightly identify this as a valid lead. This leads to a higher volume of leads and minimizes missed opportunities.

Transforming Conversations, Empowering Businesses

Live Chat stats in 2023 show that this form of customer support is the most common use for B2C businesses. However, this isn't the only use of Live Chat. It also improves CX score by identifying the pain points in customer service — **72%** of customers are satisfied or very satisfied with their **customer support experience** when shopping online, but the level of **satisfaction** increases to **92%** when it is used.

So, organizations that invest in developing an intelligent suite of chatbot ecosystem(s), where they can not only connect with the customers through a chatbot, but also mine the interactions to meet their objectives, will develop a measurable competitive advantage.

What if businesses receive real-time, direct feedback and could use that to create more customer-centric and personalized products and offers? That's the opportunity Chat Intelligence provides.

At Tiger Analytics, we've been working with clients across industries to help them translate existing data into insights, by tapping into the power of NLP and Generative AI. By leveraging **USE and BERT** embeddings, we've also taken chat analysis to new heights, improving model accuracy and business acceptance rates. Our cutting-edge deep learning based models, trained on domain and industry-specific nuances, provide a more accurate, effective, and efficient alternative to the traditional chat solutions. Through our solutions, we redefine how organizations engage with customers, making them a driving force in the journey towards **Chat Intelligence excellence.**

74%

of B2C businesses

85%

of B2B businesses use live chat for sales

31%

of B2C businesses

54%

of B2B businesses use live chat for marketing

About the Authors

Chandrachur Mukherjee

With extensive experience in Consulting, Data Science, and AI Engineering, Chandrachur Mukherjee has a proven track record of working with Fortune 500 clients across diverse industries and geographies, including CPG, Travel and Hospitality, Technology, Retail, Banking, and Financial Services.

Sabarish Gopalakrishnan

A proficient Data Scientist, specializing in supply chain and financial services, Sabarish Gopalakrishnan excels in implementing end-to-end solutions (prototyping, developing, and deploying ML models at scale) that drive efficiency and growth.

Sources:

- Live Chat Statistics Your Business Needs in 2023 - Link: https://techjury.net/blog/live-chat-statistics/
- 25 Reasons Live Chat Can Help You Grow Your Business In 2023 - Link: https://www.superoffice.com/blog/live-chat-statistics/)
- MLM — Sentence-Transformers documentation (sbert.net) - BERT MLM diagram
- https://www.cs.ubc.ca/labs/algorithms/Projects/Ablation/
- https://www.mitel.com/features-benefits/chat-logs
- https://www.jdpower.com/business/press-releases/jd-power-2016-us-chat-and-email-benchmarking-study
- https://www.cabinetm.com/product/goto/bold360
- https://www.investor.gov/additional-resources/retirement-toolkit/self-directed-plans-individual-retirement-accounts-iras
- https://www.forrester.com/blogs/retailers-without-chat-a-missed-opportunity/

About Us

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 and 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

US | UK | Canada | India | Singapore | Australia