

Artificial Intelligence & Machine Learning Architecture of Lead Generation

Shishir Sarkar

Enterprise Architect

SWZD

Pune, India

Shishir.Sarkar406@gmail.com

Abstract—Artificial Intelligence is the most sought-after technology in today's era, there are many domains where Artificial Intelligence started impacting, and one of the domains which I bring here is Digital Marketing which I specifically discussed in the field of Lead generation area.

Lead generation is a significant field that provides a taxonomy of customers in the form of Leads to the business so that they can promote their product, by identifying the correct category of the customer where Artificial Intelligence intelligence plays a vital role.

Artificial Intelligence also plays and resolves some underlying problems that are associated with Lead generation like identifying intent-based categorization, and the accuracy of firmographic and demographic data.

There are various other stages in the process of Lead generation where we can apply Artificial Intelligence, however, it's very crucial to analyse which machine learning model we need to use and more importantly how we can train the data.

Training of the model plays a vital role in bringing accuracy to the process and thus requires the proper infrastructure where we can stream the high volume of the required data and compute engine which can process and filter the data as per the needs of training the model.

Machine learning is the sort of pathway to achieve the Artificial Intelligence, machine learning encapsulates various algorithms and as per the requirement we can refer the intended algorithm and data set, where data set play a pivotal role to train the model.

Keywords—Artificial Intelligence, Machine learning, Lead generation, demographic, formographic

Date of Submission: 23-06-2024

Date of Acceptance: 03-07-2024

I. INTRODUCTION

There is significant investment happened in past year in different country like America around 5 to 8 billion, where China has invested around 1 to 2.5 billion and Europe invested 1.7 billion. Now India playing very crucial role adopting, investing and developing AI based product, where market size is around 5.2 billion in 2023.

Artificial Intelligence plays a prominent role in the Lead and demand generation business, where we need to bring more accuracy to generating the leads. The real challenge that we are solving in leads that there are many sources where we can capture the data, one of the prominent ways we use to capture the digital footprint of customers.

However capturing the digital footprint is not the easy way, as we know, there is plethora of data is generating now a day, we there are many digital platforms augmented where user is visiting and sharing their interest, to capture this high volume of data we need a proper infrastructure to handle large volume of data.

Even we need a proper process of engine which can filter and execute this volume of data, because there are many other dataset which is not needed for our AI system and machine learning model, which has to filter it out.

As we know, to train the model we need fine dataset that is we achieve through proper data processing engine.

II. PROBLEM STATEMENT

To achieve a greater level of accuracy in lead and demand generation using AI, we need to have clean and noiseless data to process using AI. Until and unless we are not able to provide clean and proper data, we won't get the desired result.

As we capture the data from different sources which is further classified into first-party and third-party data.

Mostly all this data is accumulated from the different different sources and most of the time the data is not clean and has a lot of noise and a missing footprint.

Here I am going to address the way where we can process the data with a proper ETL framework so that we can achieve the highest level of accuracy of data.

And then how these data will be further processed by AI and machine learning models to achieve the proper categorization, intent data, firmographic data, and demographic data.

III. DATA PROCESSING ENGINE

Now in today world we gathered high volume of data on daily basis and there many platform which generating high volume of digital footprint of the customer hourly basis, however most of the time this data is not cleaned. To gather and understand this data we need certain system which can consume the data and process the data so as resultant we will get the required data which can further feed to the machine leaning models to get the different taxonomy of customer data.

However to consume and process this volume od data on daily basis is not the easy task.we need certain framework and platform which can handle this data.

Generally before cloud service we were leverage using hadoop based big data platform like Cloudera platform to execute large volume of data, now a days we can refe the cloud managed services as well like AWS EMR, google dataproc and azure databrics.

However there is further deep analysis require what is the size of cluster and executors require to execute the ETL process.

Below is the set of framework which can be effective way to consume, and filter out the different set of noise in the data.

Data engine workflow

Data standardisation - Data standardisation is the process where we flatten the data in the standard format, as we aggregate the data from different sources and different aggregator agencies have their predefined format which is sometimes different from the standard format, so before proceeding with further step we need to bring it in standard formatting of the data.

Data validation - Data cleaning is a vital activity that needs to be carried out while consuming data where we can filter out empty information and null values essentially any record which doesn't follow the intended data type standard, those kinds of data are politer out during that process because those data is biggest risk to have the noise in the dataset and also augmenting the size of data in the data store which eventually needs to pay the higher costing and also decrease the efficiency of machine learning model.

Data Sampling - Data sampling is the process of carrying out the subset of the intended data as per the need. Generally, third-party aggregators feed large numbers of data, however, as per the requirement we need to get the required set of data which can suit our requirement, so in this process, we sample the subset of the data as per our requirement.

Data curing - Data curing is an important process where we can take corrective measures to enrich the data and aids in guaranteeing that data is well-organized, safe, accessible, and of excellent quality. Organizations may maximise their data and guarantee its efficient use by adhering to certain best practices.

Data Transformation - This is one of the major processes where we map the collected dataset into desired entities and attributes that could be in the unstructured format or in the structured format, this transformation generally takes a long time to execute and the resultant data of this process would be the actual intended data which further load and feed into the machine learning models.

Data loading - This is generally a final process where we load the data in the target data store, data lake, data mart or data warehouse. In this process as per the requirement, we eventually load the data incrementally, however as per the need we can load the data based on the timestamp or randomly.

The failover scenario plays a vital role here, we need to also check if any failure happens the fail over recovery option should be there, so that we can make sure data loads effectively.

Refer section X.B.a for the architecture diagram

IV. IDENTIFYING POTENTIAL CUSTOMERS THROUGH THE CATEGORISATION TECHNIQUE.

Artificial Intelligence plays a critical role in implementing categorization techniques. Categorization is a way to create a different product category and an association with the demographic data.

Identifying and creating different product categories will require proper datasets and algorithms. In our study, we employ the supervised machine learning technique to categorize possible customers. Initially, we preprocessed the data after gathering it from social media and search engines, which is a digital footprint of customer behavior. Features are defined as the number of events that happen by the customer in the intended product. To label the data, we can build the dataset. We have trained our supervised model for classification using this labeled data. Following the evaluation of a customer's purchasing behavior. Computational measures relating to classification are acquired from the approaches. Pre-processing of data is a major part of this system. total understanding of the information needed to identify the current features. To improve the accuracy, features were designed.

V. MAXIMIZE EFFECTIVENESS AND CORRECTNESS OF INETNET DATA

Using Artificial Intelligence, we analyze the intent behavior of customer which helps to know what kind of intent customer have, which helps to analyze which kind of product customer is looking for, further this helps to analyze and categorize the customer intent.

Customer intent data can be further use in multidimensional way in intent mapping, intent segmentation and intent scoring.

In all the above role of Artificial Intelligence plays crucial role in intent mapping where we can collect the intent data from the different source and using Artificial Intelligence we can map the data with the correct category.

Another dimension is Intent segmentation where Artificial Intelligence plays a critical role to extract correct intent behaviour data and analyze intent based signals from business behaviour patterns, if the segment is not exist, than analyze the new segment based on the intent behaviour pattern.

There are approached where applying Artificial Intelligence would be analyze the brands to analyze which subsegment which they expect and desire to target.

Another use case where we can apply Artificial Intelligence approach to discover hidden pattern in behaviour data and supply possible sub segment emerging from unsupervised machine learning method.

Artificial Intelligence also plays the significant role to identify the new segment and the associated subsegment.

The more important dimension is to identify the intent scoring, which helps sales and marketing end users to produce reasonable account ranking of the current week using underlying intent raw data. The account ranking produced by the score should strike a balance between adapting and recent changes in intent activity with historical weighted activity.

The score must strike a balance between tacking and smoothing with the best score responding strongly to significant activity movement.

There is quite a general formula that probably take is as a base for designing algorithms:

Let X be the vector of historical data for a given account across all the data sources. Find a function

$f: X \rightarrow y$ where y is a scaler representing the overall level of buyer intent at the current week.

F must satisfy the requirement listed below in section B.

The ultimate choice of f will be the one that best meets the success criteria as mentioned below

Measure the quality of the score and algorithm judged by the combination of criteria like tracking, smoothing, company size, spread, and boundness.

Refer section X.B.b for the architecture diagram

VI. INTENT BASED ENGINE

Using Big Data and computational calculations, predictive analytics, also known as predictive intent, forecasts who will buy what and when. Data is gathered from a variety of data sources, including sales and customer profiles. Data mining, data modeling, machine learning models, AI, and historical data are all included. It uses historical data—rather than real-time browsing or purchasing activity—and human expertise—which is rather limited—to determine the most effective way to reach a particular audience.

On the other hand, buyer intent data is gathered through monitoring and documenting real customers' online activities. Intent data is more reflective of actual activity, so you can use it to help sales representatives and marketers find new clients by using relevant content.

An accurate analysis of a user's search intent shows what they are looking for when they type a query; it's highly targeted information without being obtrusive. An intelligent search engine will be able to determine the user's purpose from the text they enter in the search box and utilize machine learning techniques to rank results as highly as possible.

Intelligent search, like other computer science-related fields, incorporates machine learning and natural language processing. It can, for instance, draw connections between semantic phrases that would be impossible for a conventional search engine to understand by just looking at keywords. The searcher intends that organizations need to take note of, recognize, and appropriately interpret. Regarding the retention of customers

Refer section X.B.b for the architecture diagram

VII. CONCLUSION

Artificial Intelligence (AI) describes methods that enable machines to carry out cognitive tasks that call for human intelligence. They consist of reasoning, learning, and interacting with the environment around the machine. Among the most well-known AI methods are machine learning and deep learning. AI can personalize brand experiences, which facilitates the development of user engagement and loyalty. Artificial intelligence based on language is developing quickly. It "learns" from past mistakes and automatically improves to produce even better results the next time. It can help marketers and companies find better leads of customer demographics and firmographic info so that they can reach out to them for better leads to convert prospects into target customers.

Through this approach, we can achieve insightful information for businesses looking to boost sales performance and strengthen relationships with customers through data-driven customer profiling. The wide range of consumer demands, behaviours, and preferences shown in online business platforms facilitates the raw data and collects the various taxonomies of customer data using the aforementioned data engine process and artificial intelligence technique.

VIII. Practical Implication

In our practise we applied above mentioned research and implemented a solution based on AI, machine learning model and data processing engine, we observed that we are able to filter different taxonomy of customer data, like demographic, firmographic and discover the different pattern of categorization and intent data behaviour.

Through this research we able to discover and generate the lead data almost close 70 to 80%. Yet there are unknown area where we need to do more research to achieve higher accuracy.

IX. Future scope

While working in this research and practical implication, we came across few of the problems which is need to research it our further, one of the problem which I want to discuss it here.

There are few pattern we observed that when process the aggregated data than we process and extract the demographic data and firmographic data, demographic data are not much dynamic in nature, however firmographic data is timebased data where its keep changing based on real world factor, like when person changes the organisation, designation, phone number or place. The problem is to verify this one it seems difficult because volume of data is so huge that we can;t use any social media or search engine platform to check and update otherwise it will endup with paying huge amount and technically there would be chances the data become inconsistent.

Declaration

Conflict of interest The authors declare no conflicts of interest related to this work.

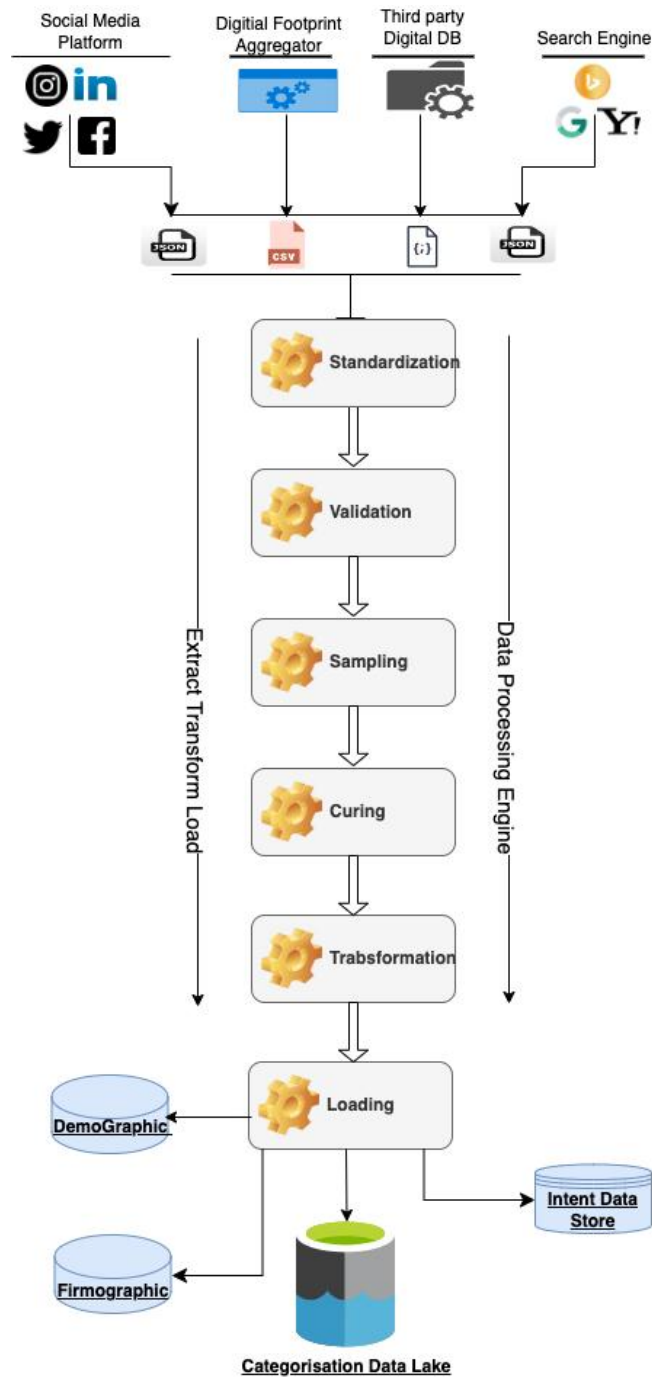
A. Abbreviations and Acronyms

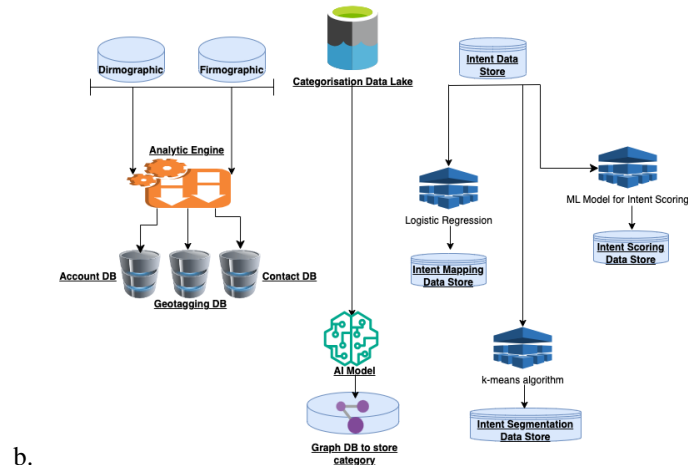
AI Artificial Intelligence

ML Machine Learning

ETL Extract Transform Load

B. Figures and Tables





b.

REFERENCES

- [1] S. Kuchlous and M. Kadaba, "Short Text Intent Classification for Conversational Agents," 2020 IEEE 17th India Counc. Int. Conf. INDICON 2020, 2020, doi: 10.1109/INDICON49873.2020.9342516.
- [2] J. Schuurmans and F. Frasincar, "Intent Classification for Dialogue Utterances," IEEE Intell. Syst., vol. 35, no. 1, pp. 82–88, 2020, doi: 10.1109/MIS.2019.2954966.
- [3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [4] A. Patra and D. Singh, "A Survey Report on Text Classification with Different Term Weighing Methods and Comparison between Classification Algorithms," Int. J. Comput. Appl., vol. 75, no. 7, pp. 14–18, 2013, doi: 10.5120/13122-0472.
- [5] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. From data mining to knowledge discovery in databases Artificial Intelligence Magazine, (1996), 37-53.
- [6] Han, J., and Fu, Y. Discovery of multiple-level association rules from large databases. Proc. International Conference Very Large Databases, Zurich, (1995).
- [7] Alsayat A (2023) Customer decision-making analysis based on big social data using machine learning: a case study of hotels in mecca. Neural Comput Appl 35:4701–4722
- [8] Das S, Nayak J (2021) Customer segmentation via data mining techniques: state-of-the-art review. Comput Intell Data Min: Proc ICCIDM 2022:489–507
- [9] T. E. Lin and H. Xu, "A post-processing method for detecting unknown intent of dialogue system via pre-trained deep neural network classifier," Knowledge-Based Syst., no. August 2019, 2019, doi: 10.1016/j.knosys.2019.104979.
- [10] Gronroos, Christian. "Relationship approach to marketing in service contexts: The marketing and organizational behavior interface." Journal of business research 20, no. 1, pp. 3-11, (1990).
- [11] Graves, Alex, Abdel-rahman Mohamed, and Geoffrey Hinton. "Speech recognition with deep recurrent neural networks." In Acoustics, speech and signal processing (icassp), 2013 IEEE international conference on, pp. 6645-6649. IEEE, 2013.
- [12] Graves, Alex, Abdel-rahman Mohamed, and Geoffrey Hinton. "Speech recognition with deep recurrent neural networks." In Acoustics, speech and signal processing (icassp), 2013 IEEE international conference on, pp. 6645-6649. IEEE, 2013.
- [13] Kashwan KR, Velu C (2013) Customer segmentation using clustering and data mining techniques. Int J Comput Theory Eng 5:856
- [14] Brito PQ, Soares C, Almeida S, Monte A, Byvoet M (2015) Customer segmentation in a large database of an online customized fashion business. Robot Comput-Integr Manuf 36:93–100
- [15] He X, Li C (2016) The research and application of customer segmentation on e-commerce websites. In: 2016 6th International Conference on Digital Home (ICDH), IEEE, pp 203–208