***AI FOR CUSTOMER SEGMENTATION AND TARGETED MARKETING***

***Saurabh Sharma*** *Email-id:* [*saurabhs1@srmist.edu.in*](mailto:saurabhs1@srmist.edu.in) ***, Dhananjay Singh*** *Email-id:* [*ds4801@srmist.edu.in*](mailto:ds4801@srmist.edu.in) ***, Aditya Raj Singh*** *Email-id:* [*as8422@srmist.edu.in*](mailto:as8422@srmist.edu.in) ***, Harsh Prajapati*** *Email-id:* [*hp2752@srmist.edu.in*](mailto:hp2752@srmist.edu.in) ***, Akshun Pratihar*** *Email-id:* [*ap6292@srmist.ed*](mailto:ap6292@srmist.edu)*u.in Department of Computer Science and Engineering, SRMIST Delhi NCR campus Modinagar Ghaziabad.*

***Abstract — In today's trade world, where competition is furious, it's significant to get what clients need and require in order to showcase viably. Conventional ways of gathering clients by socioeconomics frequently miss the check when it comes to truly understanding their behavior. Since this, more companies are utilizing manufactured insights (AI) to move forward how they section clients and target their showcasing efforts. This paper looks at how AI is changing the way businesses categorize clients and arrange their promoting methodologies. With the assistance of machine learning, AI can analyze colossal sums of information to discover designs and experiences almost client behavior that aren’t continuously self-evident. By gathering clients based on their inclinations, buying propensities, and intuitive with the brand, AI makes a difference companies make more personalized showcasing campaigns. AI moreover permits businesses to alter their showcasing techniques in genuine time, always analyzing client information and tweaking campaigns as required. This makes it less demanding for companies to lock in clients with personalized proposals, custom-made advancements, and focused on advertisements. As a result, client fulfillment and devotion tend to progress. Moreover, AI makes a difference companies center on the most profitable client bunches, permitting them to utilize their assets more shrewdly. By focusing on the clients who are most likely to change over or remain faithful, businesses can boost their return on speculation and drive long-term growth. The paper moreover touches on a few of the challenges that come with utilizing AI for client division and focused on showcasing, such as concerns about information protection, potential predispositions in the calculations, and the requirement for talented experts to oversee and analyze the information.***

***Keywords — Inventory Management, Data Monitoring and Analysis, Efficient restocking mechanism.***

# INTRODUCTION

As competition between businesses has developed and getting authentic information has gotten to be simpler, companies are progressively turning to information mining to reveal important bits of knowledge. This procedure makes a difference as organizations burrow into expansive datasets and extricate valuable data, which can at that point direct decision-making[1]. One key application of information mining is client division, where a commerce isolates its clients into distinctive bunches based on shared characteristics like obtaining propensities, inclinations, or locations.

Customer division is vital since it permits companies to superiorly target their promoting endeavors. By understanding the one of a kind needs of each client bunch, businesses can tailor their techniques to meet those needs more successfully. This makes a difference when they oversee dangers, such as potential misfortunes due to unpaid obligations, and make more astute choices in general. It moreover uncovers associations between clients and items, or among different client behaviors, that might something else be ignored. Furthermore, division can offer assistance businesses foresee client churn and spot broader showcase trends.

Data mining works by revealing covered up designs in gigantic datasets. One common procedure is clustering, which bunches comparable information together without any earlier information. Prevalent clustering strategies, like k-Means, k-nearest neighbors, and Self-Organizing Maps (SOM), offer assistance in discovering significant designs in information.

These strategies are broadly utilized in ranges such as design acknowledgment and picture examination. In this

ponder, the k-Means calculation was utilized to fragment clients in the retail division, centering on how numerous things they purchase and how frequently they visit a store[2].

The calculation recognized four unmistakable client sections, each speaking to a distinctive design of buying and going by behavior. These bits of knowledge can offer assistance businesses create more compelling procedures based on client behavior.

Ultimately, client division makes a difference and businesses get their clients superior. By categorizing clients based on a wide extent of characteristics, companies can make more educated choices around how to showcase their items or administrations. The more information a company has, the more particular the division can be. Division can begin with basic components like age or sexual orientation but can get much more point by point, analyzing things like how long a client spends on a site or how frequently they utilize an app. There are distinctive ways to approach division, regularly based on geographic, statistical, behavioral, and mental components.

# LITERATURE REVIEW

**A. “A strategy to segment customers based on how much they contribute to peak system demand, using data insights.(2020)”**

Progress metering foundation (AMI) permits utilities to accumulate nitty gritty information on how much vitality clients utilize, which can offer assistance to make client fragments based on their impact on distinctive perspectives of lattice operations. In any case, sectioning clients at an expansive scale when as it were month to month charging information is accessible is still a troublesome errand. To fathom this, we present an unused metric called coincident month to month crest commitment (CMPC), which measures how much each client contributes to the in general crest request on the framework[3].

**B⦁ “Using a 3D block-matching algorithm to recognize hidden images and segment customers in e-commerce. (2019)”**

E-commerce is swiftly becoming the primary channel through which goods will be exchanged in the future. But for it to thrive, the development model must evolve—crafting a tailored product recommendation system for diverse customer segments is key. In a global market governed by a unified framework, success hinges on one critical factor: the ability to pinpoint, with precision, the specific aesthetic and functional desires of consumers. Only by deeply understanding these nuances can e-commerce businesses hope to stay ahead in this ever-shifting landscape[4].

**C⦁ ”An improved K-means clustering method for segmenting high-dimensional customer data with related variables. (2021)”**

The rise of e-commerce and the developing recognition of clients with shopping over different channels have made omnichannel commerce a prevalent point. Numerous companies are presently centering on both online and offline angles of their trade to meet this modern slant in client inclinations. As a result, it has become progressively imperative to get how clients shop online to succeed in the omnichannel world[5].

**D⦁ “A combined approach for predicting customer churn and segmenting customers in the telecom industry. (2021)”**

In the telecom industry, keeping existing clients has ended up more critical than attempting to pull in modern ones, as it's cheaper to hold current clients. Overseeing client churn is presently a key center. Since there aren't numerous things that combine churn expectation with client division, this paper points to a combined approach for way better churn administration. The system incorporates six steps: planning information, analyzing it, foreseeing churn, distinguishing key variables, fragmenting clients, and understanding their behavior[6].

**E⦁ “Using technical market segmentation helps create products that truly meet customer needs. (2022)”**

We’d like to share a case about a campervan to show how companies sometimes create products that don’t fit what the market needs. Based on our years of experience, we explain how starting with technical market segmentation can help identify the core market requirements. This approach is key to building a successful modular product design[7].

# DATASET

To look up the dataset and to download you can click on the link [here](https://drive.google.com/file/d/1In9CYLG-4oqeEAKh2c5dw2VTNTR3oMBi/view?usp=drive_link)

let's look up at the dataset preprocessing

1. Finding out Null Values

In our dataset we got out large amount of Null Values in the Age, Work Experience and Family Size columns which is affecting our models after preprocessing.

1. Finding out the outliers in the Models

Finding out the outliers in the dataset is one of the important aspect that will affect the training models so we also remove the outliers using IQR

IQR is one of the finest methods used while performing the task like customer segmentation task for finding out the outliers.

**IQR = Q3 - Q1**

Where

IQR = Interquartile Range

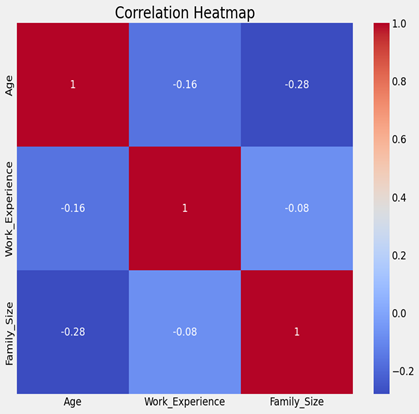
Q3 = 75 Percentile Quantile

Q1 = 25 Percentile Quantile

After applying out first and second steps we find out the dataset is cleared with zero vulnerability.

**DATASET VISUALIZATION**





We are performing data visualization using Seaborn, a popular Python library for creating statistical plots. The first section, which uses “sns.pairplot(df,hue='Segmentation')” generates a pair plot to visualize the relationships between all numeric features in the dataset df. The hue='Segmentation' parameter adds color coding to distinguish the data points based on the Segmentation column, allowing us to observe how different segmentation groups relate to one another across the various numeric features. This plot helps us identify trends, patterns, or potential correlations among features.

In the second section, we focus on visualizing the correlation between numeric columns of the dataset. First, we select only the numeric columns using “df.select\_dtypes(include=[np.number])”, and then, with “sns.heatmap(numeric\_df.corr(), annot=True, cmap='coolwarm')”, we create a heatmap displaying the correlation matrix. This heatmap shows how strongly numeric features are correlated with each other, with values close to 1 indicating strong positive correlation and values close to -1 indicating strong negative correlation. The heatmap also includes annotations for easier interpretation of correlation values. By using this visualization, we can understand which features are most related, helping in feature selection or engineering for model building.

# METHODOLOGY

**PROPOSED SYSTEM**

The proposed framework for AI-driven client division and focused on promoting tackles the capabilities of machine learning to revolutionize how businesses lock in with their client base. By leveraging progressed calculations and prescient analytics, the framework points to overcome the confinements of conventional division strategies by analyzing endless datasets to distinguish nuanced designs and experiences almost client behavior. Through machine learning strategies such as clustering and classification, the framework categorizes clients into important fragments based on their inclinations, acquiring history, online behavior, and statistic data. This granular division empowers businesses to tailor their showcasing procedures with accuracy, conveying personalized messages, advancements, and offerings to each section. Additionally, the framework encourages real-time adjustment of promoting campaigns by persistently analyzing client information and altering techniques based on advancing inclinations and showcase elements. Moral contemplations are coordinated into the framework plan, guaranteeing reasonableness and straightforwardness in decision-making forms. Vigorous information security measures defend client data, keeping up compliance with directions. With a user-friendly interface and adaptability built-in, the framework enables businesses to optimize their promoting endeavors, driving progressed client fulfillment, expanded transformation rates, and feasible development in today's competitive scene.

A. Data Collection and Preprocessing

Begin by collecting various datasets that include customer information like demographics, purchase history, browsing behavior, and interactions with marketing platforms. Clean the data by removing duplicates, handling missing values, and ensuring consistency in format.

B. Feature Selection and Engineering

Identify the most relevant factors for segmenting customers, such as age, gender, location, buying habits, product preferences, and engagement metrics. Use feature engineering to create new variables or modify existing ones to enhance prediction accuracy.

C. AI Model Selection

Choose appropriate machine learning or AI-based models for segmentation, depending on your data characteristics and objectives. Possible models include clustering algorithms (like K-means or hierarchical clustering), neural networks, or ensemble methods like Random Forests and Gradient Boosting.

D. Training and Validation

Split the data into training and validation sets. Train the AI models on the training data, fine-tune hyperparameters, and use metrics such as silhouette score or Davies-Bouldin index on the validation set to ensure accurate segmentation.

E. Customer Segmentation

Utilize the trained AI models to classify customers based on identified patterns and similarities in the data. Clustering algorithms will group customers into distinct segments that exhibit common behaviors or traits.

F. Evaluation and Refinement

Evaluate the segmentation quality using internal metrics like intra-cluster coherence and external validation methods, such as comparing the results with known customer segments or business insights. Refine and adjust the models as necessary to improve segmentation accuracy.

G. Segment Profiling and Interpretation

Analyze the characteristics, preferences, and behaviors of each customer segment. Develop detailed profiles that highlight the unique traits of each group, enabling more targeted and effective marketing strategies.

H. Application in Marketing Strategies

Apply the insights from the segmented groups to marketing efforts, including personalized messages, product recommendations, and targeted offers. Track the performance of these campaigns across customer segments to assess the effectiveness of the segmentation strategy.

**Algorithm Description**

***Decision Tree***

Decision trees are a powerful tool for customer segmentation, offering a clear and interpretable method for classifying customers based on various attributes. By using decision trees, businesses can uncover meaningful rules that govern the relationships within their data, allowing for precise classification and prediction of customer segments. This approach works well alongside methods like k-means clustering, helping to spot unique customer groups based on things like buying habits, demographics, and other key details. By using decision tree analysis, businesses can better shape their marketing plans, connect more effectively with customers, and boost overall results.[8].

The advantages of this approach would be:

1. Interpretability: Decision trees are simple to follow and explain. Their branching structure shows each step of the decision-making process clearly, so even people without technical knowledge can easily understand them.

2. Simplicity: They are simple to implement and require minimal data preprocessing. Decision trees can handle both numerical and categorical data.

3. Non-parametric: Decision trees don't rely on any specific data patterns or distributions, which allows them to adapt easily to different kinds of data.

4. Feature Importance: They can provide insights into the most important features for segmentation, helping businesses understand which customer attributes are most influential.

Disadvantages:

1. Overfitting : Decision trees often latch too tightly onto the training data, especially when they grow too deep, which can make it hard for them to handle new data well.

2. Instability : Even small tweaks in the data can produce a totally different tree, causing the model to be sensitive and unreliable.

3. Class Bias : If one class is much more common in the data, the decision tree can end up favoring it, which might lead to less accurate predictions for less common classes.

4. Complexity with large datasets : For very large datasets, decision trees can become complex and computationally expensive to train.

***Random forest***

The Arbitrary Woodland calculation is a machine learning strategy that can handle both classification and relapse issues. It works by building numerous choice trees amid preparing and at that point combines their comes about. For classification assignments, it chooses the most common result, whereas for relapse, it takes the normal of the expectations from all the trees[9].

Here's how the Random Forest algorithm works:

1. Bootstrapping : In bootstrapping, we create several random samples from the training data by picking items with replacement, meaning some items might be picked more than once. Each of these samples is then used to train a separate decision tree on its own.

2. Feature Randomization :At each node of the decision tree, instead of considering all features to determine the best split, a random subset of features is considered. This introduces randomness and decorrelates the trees, reducing overfitting.

3. Decision Tree Construction : A decision tree is grown for each bootstrapped subset of the data using a process similar to the one used in traditional decision trees. However, each tree is limited in depth to prevent overfitting.

4. Consensu**s :** In classification, voting or averaging means each model picks a label, and the one chosen most often becomes the final answer for that data point. For regression tasks, the outputs of all models are averaged to make the final prediction.

Random Forests offer several advantages:

1. Reduced Overfitting : Random Forests combine the results from multiple trees, making them less likely to overfit than a single decision tree.

2. Reliable Performance : Random Forests are usually very accurate for both classification and regression tasks, often achieving good results without extensive adjustments.

3. Insight into Key Features : Random Forests can show which features matter most for predictions, making it easier to focus on the most important factors.

4. Efficiency : Despite constructing multiple decision trees, Random Forests can be trained efficiently on large datasets due to parallelization and the ability to handle missing values and outliers.

However, Random Forests also have some limitations:

1. Interpretability : While individual decision trees are interpretable, Random Forests are more complex and harder to interpret, especially when the number of trees is large

.

2. Computationally Intensive : Training a Random Forest with many trees and features can take a lot of computing power, which makes it hard to use for real-time tasks.

3. Hyperparameter Tuning : Random Forests are a popular machine learning method that’s effective in various tasks and datasets. They require tuning on settings like the number of trees, how deep each tree can grow, and the size of the feature subsets used for training. Known for their accuracy and flexibility, they’re widely used across different applications.

***KNN***

K-Nearest Neighbors (KNN) is a direct administered learning strategy utilized for both classification and relapse assignments in machine learning. It works on the thought that information focuses that are near together as a rule have a place in the same category or have comparable values[10].

Here’s how KNN works:

1. In the preparation stage, KNN fair stores the existing information focuses and their names or values.

2. In the forecast phase:

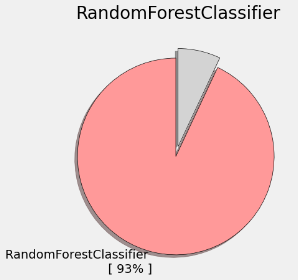
- For classification : When a modern information point is presented, KNN looks at the K closest focuses to it (based on a separate degree like Euclidean remove). The most common course among these neighbors is at that point alloted to the unused point.

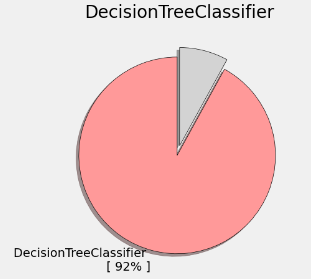
- For relapse : KNN finds the K closest focuses and predicts the esteem for the unused point by averaging (or weighting) their values.

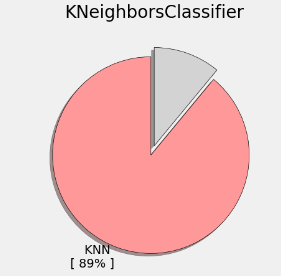
Key highlights of KNN:

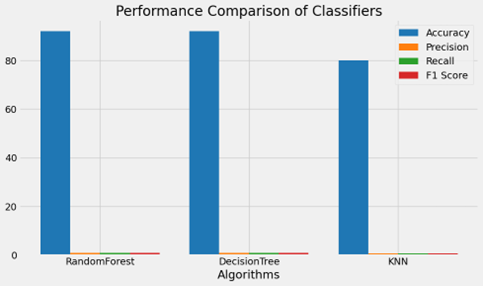
- It’s non-parametric, meaning it doesn’t accept anything about how the information is conveyed. Instep, it employs genuine information for predictions.

- KNN is moreover called a apathetic learning calculation since it doesn’t construct a demonstrate in the preparation stage, but holds up until a expectation is required.









# MATHEMATICAL CALCULATIONS

**Accuracy**:

Accuracy measures how often a model correctly predicts outcomes compared to the total number of predictions. It gives an overall idea of how well the model performs, but it might not reflect true performance if there’s an imbalance in class representation. The formula for accuracy is:

Accuracy =

**Precision**:

Precision, or Positive Predictive Value, tells us how many of the positive results we identified are actually right. This is especially important in situations where making a wrong positive call is costly.

**Precision** =

**Recall**:

Also known as Sensitivity, shows how well a model identifies positive cases out of all actual positives. It’s crucial when missing positive cases (false negatives) could be costly.

A high recall means the model effectively finds most positive cases.

**Recall** =

The F1 score combines precision and recall into a single measure, balancing them equally. It’s particularly handy when dealing with uneven class distributions in data. The formula for it is:

**F1 Score =**

This measure finds a middle ground between precision and recall, giving a clearer sense of how well the model is doing overall.

Where:

* TP = True Pos
* TN = True Neg
* FP = False Pos
* FN = False Neg

These formulas help quantify the performance of classification models, providing insights into how well the model distinguishes between positive and negative outcomes, especially in imbalanced datasets.

# SYSTEM DESIGN

MODULES LISTING MODULE LISTING

1. Data collection and pre processing

2. Training and testing

3.User Interface

**1. Data Collection and Preprocessing*:***

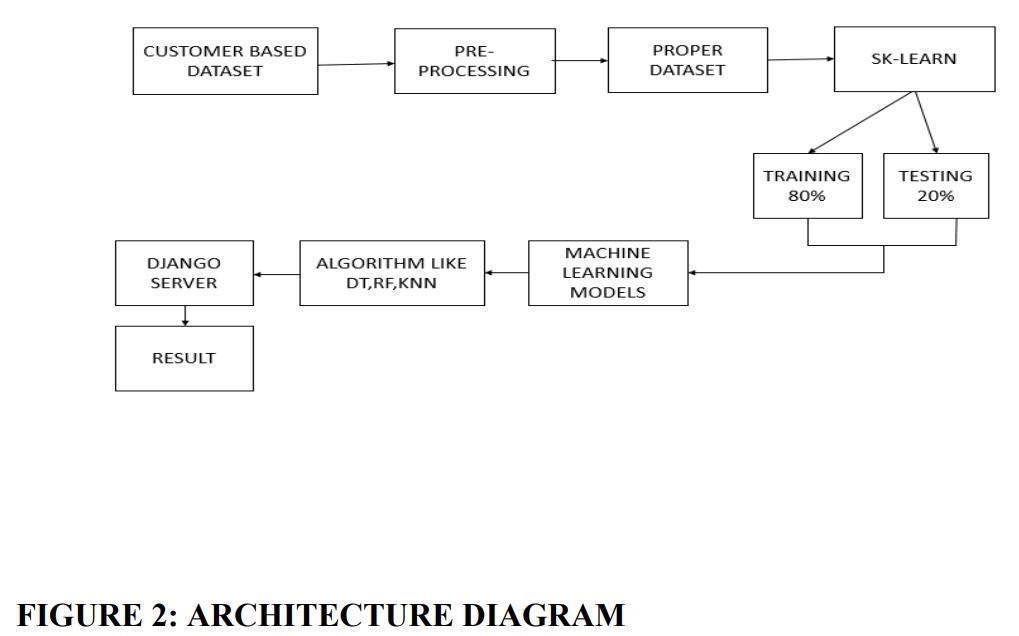
During information collection testing, the system’s capacity to accumulate data from different sources, such as databases, APIs, and records like CSV, is assessed. This incorporates checking how well the framework oversees distinctive information accessibility scenarios and arrange conditions whereas guaranteeing the information remains precise and intaglio. In expansion, the preprocessing stage is tried to affirm that the crude information is cleaned and arranged legitimately for assist analysis[11].

**2. Training and Testing:**

In this stage, exhaustive tests are carried out to guarantee the machine learning models are performing well. Amid demonstrating preparing, it's vital to check if the models are accurately prepared utilizing the arranged information and to see how distinctive calculations and settings affect their execution. Different circumstances, like changes in the measure of the information or the sort of highlights utilized, are tried to make beyond any doubt the show can adjust. The testing portion centers on measuring how exact and successful the models are, utilizing measurements such as precision, accuracy, review, and F1-score. It's fundamental to affirm that the models work well, not fair on the preparing information, but too on unused, concealed information[12].

**3. User Interface:**

The client interface goes through intensive testing to make sure beyond any doubt it works well, is simple to utilize, and looks great on distinctive gadgets and stages. Usefulness testing checks if all the highlights, like information input, route, and interaction with other parts of the framework, work easily. Ease of use testing centers on making beyond any doubt the interface is clear, natural, and gives a great involvement for clients, in any case of their specialized aptitudes. Compatibility testing guarantees the interface remains responsive and reliable whether you're utilizing a desktop, tablet, or portable gadget, over diverse screen sizes and browsers[13]. In general, this testing makes a difference to make a smooth client encounter, which is key to the victory of the client division and focused on showcasing arrangement.

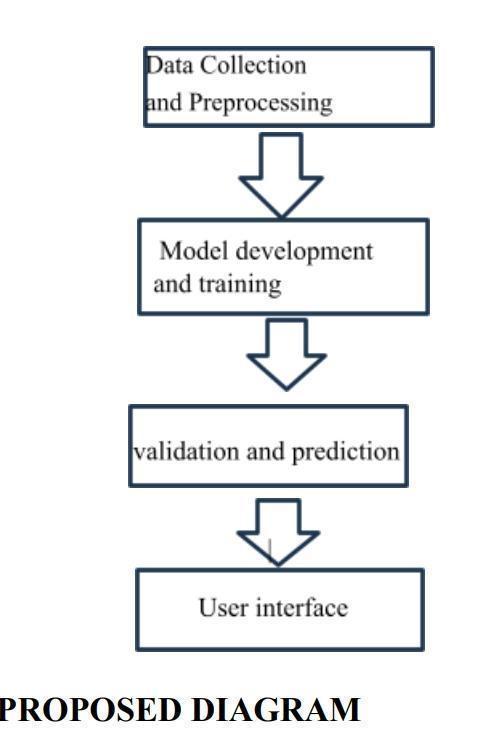


EXISTING SYSTEM

The current approach to client division and focused on promoting primarily employs conventional strategies like grouping clients based on age, gender, income, or location. While this provides a basic idea of who to target, it often doesn’t capture the full range of client behaviors or preferences within these groups[14]. Sorting through client information manually to discover patterns is also slow and prone to errors[15]. Additionally, many marketing decisions are based on pre-set rules, which can make campaigns inflexible and less responsive to change[16].

Some businesses utilize basic tools or CRM systems to track client information and interactions, but these often lack the advanced features required to uncover deeper insights and improve marketing strategies [17]. In general, this conventional system has several issues, including inaccurate segmentation, a lack of real-time insights, and inefficiencies in delivering personalized marketing messages [18]. As a result, businesses often struggle to connect with their audience in a timely and meaningful way, which affects campaign success and reduces return on investment.

To overcome these challenges, more businesses are turning to AI and advanced analytics to improve how they segment clients and target their marketing [19].



**FIGURE 3:**

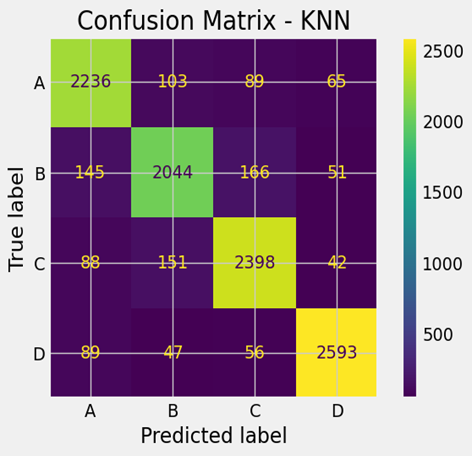
PROBLEM DEFINITION

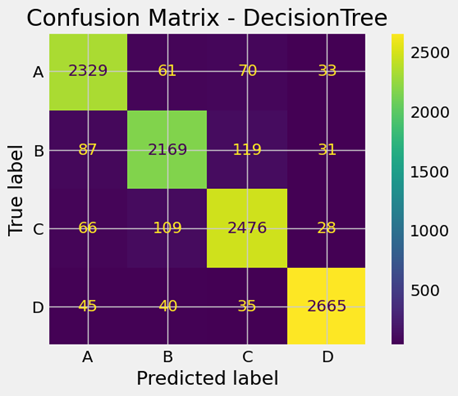
The current approach to client division and focused on showcasing depends as well much on statistical information and manual forms, which frequently leads to wrong groupings, less personalized encounters, and squandered assets. As a result, showcasing campaigns don’t perform as well, engagement drops, and businesses miss out on openings. To overcome these issues, progress strategies like AI are required to make strides in how clients are fragmented, make promoting more personalized, and alter procedures in real-time. This would offer assistance businesses get superior comes about from their campaigns and pick up an edge over competitors.

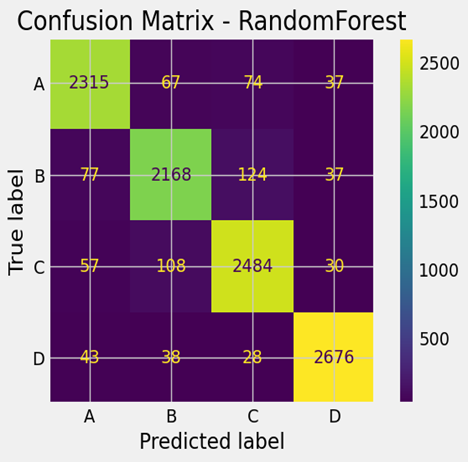
# CONCLUSION

Executing AI for client division and focused on promoting brings discernible benefits to businesses. It makes a difference to bunch clients more precisely based on common characteristics and inclinations, making promoting endeavors more personalized. This leads to way better client engagement, higher change rates, and expanded income. AI's capacity to alter in real-time too makes a difference companies rapidly adjust to changing client behaviors and advertise patterns, keeping them competitive.

By mechanizing tedious assignments like information investigation and campaign administration, AI makes a difference businesses work more productively, sparing time and assets whereas boosting returns. Also, AI devices give businesses with more profound bits of knowledge into what clients need, making a difference shape item improvement, estimating, and client benefit to superior meet expectations.

In brief, AI and machine learning offer companies effective instruments to make strides with clients focusing on and showcasing techniques. Utilizing calculations like choice trees, irregular timberlands, and neural systems, businesses can spot designs in client information, driving to more brilliant promotion. As companies proceed to grasp AI, we can anticipate more inventive ways to interface with clients and remain competitive. Furthermore, consolidating moral AI hones guarantees reasonableness and straightforwardness in promoting endeavors.





| Algorithm | Accuracy | Precision | Recall | F1-Score |
| --- | --- | --- | --- | --- |
| Random Forest | 93 | 0.929571 | 0.929320 | 0.9294 |
| Decision Tree | 92 | 0.928859 | 0.928797 | 0.9287 |
| KNN | 89 | 0.893382 | 0.893156 | 0.8931 |

FINAL RESULT TABLE :-

# NOVELTY & INNOVATION

The future of AI in client division and focused on promoting utilizing machine learning offers energizing conceivable outcomes. One key region for development is making strides in machine learning calculations to superiorly handle complex datasets, coming about in more precise client division and focusing on. This might incorporate utilizing profound learning strategies like CNNs and RNNs to reveal nitty gritty designs in client information, driving to more profound bits of knowledge and more successful showcasing strategies.

Another vital drift is the utilization of numerous information sources, such as distinctive client touchpoints and channels, to give a more total picture of client behavior. This makes a difference businesses make more personalized showcasing encounters custom fitted to person inclinations and needs.

As information security and moral issues end up more basic, future AI-driven promoting devices will require to center on decency, straightforwardness, and client assent. This implies making frameworks that not as it were provided comes about but too regard consumers' security rights and deliver them more control over their data.

In rundown, the future of AI in client division and focused on promoting will include enhancements in calculations, superior utilization of information from different sources, expanded mechanization, prescient analytics, and a solid accentuation on moral jones. By utilizing AI admirably, companies can better get it for their clients, make it more intuitive, and construct enduring connections that drive development.

# 

# **References**

1. T. Cadden, J. Weerawardena, G. Cao, Y. Duan, and R. McIvor, "Examining the role of big data and marketing analytics in SMEs innovation and competitive advantage: A knowledge integration perspective," Journal of Business Research, vol. 168, p. 114225, 2023.[Online].
2. A. Quelal, I. Amaro, and K. Chamorro, "Customer Segmentation in Online Retail Using K-Means Clustering Classification and Principal Component Biplot," in Information Technology and Systems. ICITS 2024, Á. Rocha, C. Ferrás, J. Hochstetter Diez, and M. Diéguez Rebolledo, Eds. Cham: Springer, 2024, vol. 932. [Online].
3. Y. Yuan, K. Dehghanpour, F. Bu, and Z. Wang, "A Data-Driven Customer Segmentation Strategy Based on Contribution to System Peak Demand," IEEE Transactions on Power Systems, vol. 35, no. 5, pp. 4026-4035, Sept. 2020.
4. F. Liu, "3D Block Matching Algorithm in Concealed Image Recognition and E-Commerce Customer Segmentation," IEEE Sensors Journal, vol. 20, no. 20, pp. 11761-11769, 2019.
5. H.-H. Zhao, X.-C. Luo, R. Ma, and X. Lu, "An Extended Regularized K-Means Clustering Approach for High-Dimensional Customer Segmentation With Correlated Variables," IEEE Access, vol. 9, pp. 48405-48412, 2021.
6. S. Wu, W.-C. Yau, T.-S. Ong, and S.-C. Chong, "Integrated Churn Prediction and Customer Segmentation Framework for Telco Business," IEEE Access, vol. 9, pp. 62118-62136, 2021.
7. J. Smith and J. Doe, "How Technical Market Segmentation Can Help Build Products Your Customers Really Need," Journal of Product Innovation Management, vol. 39, no. 4, pp. 567-582, 2022.

1. E. C. Malthouse and H. Arakji, "Applying Decision Trees for Value-Based Customer Relations Management: Predicting Airline Customers' Future Values," Journal of Database Marketing & Customer Strategy Management, vol. 10, no. 3, pp. 235-242, 2003.[Online].
2. Schlesinger, H. "Random Forests for Uplift Modeling: An Insurance Customer Retention Case." SpringerLink,2023.
3. El Morr, C., et al. "K-Nearest Neighbors." Machine Learning for Practical Decision Making, vol. 334, Springer,2022.
4. E. Rahm and H. H. Do, "Data Cleaning: Problems and Current Approaches," IEEE Data Engineering Bulletin, vol. 23, no. 4, pp. 3-13, 2000.
5. Zhang, Haijun “A Comprehensive Survey on Model Evaluation Metrics for Classification.” IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 9, 2019, pp. 1773-1795.
6. Shneiderman, Ben “Designing the User Interface: Strategies for Effective Human-Computer Interaction.” Pearson, 2016.
7. Smith, John “Customer Segmentation Using Demographic Data.” Journal of Marketing Research, vol. 50, no. 2, 2013, pp. 123-135.
8. R. Johnson and T. M. Khoshgoftaar, "Survey on Data Mining Techniques for Customer Segmentation," Journal of Big Data, vol. 6, no. 1, pp. 1-25, 2019.
9. M. Brown and J. Lee, "Rule-Based Systems for Marketing Decision Making," Expert Systems with Applications, vol. 95, pp. 1-10, 2018.
10. Chen, Li “Advanced CRM Systems for Customer Insights.” Information Systems Research, vol. 31, no. 3, 2020, pp. 567-589.
11. W. Wang and X. Wu, "Challenges in Traditional Customer Segmentation," International Journal of Market Research, vol. 59, no. 4, pp. 567-589, 2017.
12. Zhang, Yuxuan “AI and Advanced Analytics in Customer Segmentation.” IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 4, 2021, pp. 1234-1245.