iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70532

A Study onCustomer Segmentation Using MachineLearning at TVS Sai Hemanth Motors, Srikalahasthi

B.Akash[1], Dr C.Nadhamuni Reddy[2]
[1]*Student, IIMBA, Department of MBA, Annamacharya Institute of Technology & Sciences::Tirupati, 9390477712,*
[2]*Principal, AnnamacharyaInstituteofTechnology&Sciences::Tirupati,9000002351,*

Abstract: In today's competitive market, understanding customer diversity is essential for tailoring marketing strategies and improving sales. Thisstudyfocusesoncustomersegmentationat TVS Sai Hemanth Motors, Srikalahasthi, using K-Means clustering, a machine learning technique.The aim is to classify customers based on demographic and behavioralfactorsandidentifymeaningfulpatternsin purchasing and non-purchasing behavior. Data was collected from both buyers and non-buyers, and the ElbowMethod wasappliedtodeterminetheoptimal numberofsegments. The analysisresultedinclearly defined clusters that revealed specific customer characteristics and preferences. This enables the dealer to make informed decisions regarding marketing strategies and customer relationship management. The study not only enhances understanding of customer profiles but also demonstratesthepracticalapplicationofdata-driven segmentation in a real-world context

Keywords: K-means analysis, Elbow method, Customer segmentation, Demographic variables

I. INTRODUCTION

Customer segmentation is a critical aspect of marketing strategy in today's competitive business environment. It involves dividing a heterogeneous customer base into smaller, more homogeneousgroupsbasedonsharedcharacteristics, suchasdemographics, behavior,andpreferences.The goalofcustomersegmentationistoenablebusinesses to target specific customer groups more effectively, thereby optimizing marketing efforts and improving customer satisfaction. In this context, businesses use various data analytics and machine learning techniques, such as K-Means clustering, to identify distinct customer segments and develop tailored strategiesthatenhancebothcustomeracquisitionand retention.

II. REVIEW OF LITERATURE

1) Priya and Sharma (2019) This research applied machine learning techniques for customer segmentation in the automobile industry. The study showed how clustering customers based on demographic and behavioral attributes helps improve lead conversion rates.
2) Naveen and Srivastava (2021) The authors examined how customer segmentation using unsupervised machine learning can help dealerships increase efficiency in customer targeting. They concluded that segmenting both buyers and non-buyers offered deeper insights.
3) Kavitha and Anil (2018) This paper discussed how data mining techniques, particularly clustering, can be used for strategic decision-making in automotive sales. The study highlighted that integrating behavioralvariablesenhancestheaccuracyof segments.
4) RajeshandMeena(2022)Thestudyutilized K-Means clustering to analyze consumer behavior in two-wheeler dealerships. It suggested that recognizing different clusters like price-sensitive, brand-conscious, and first-time buyers helps in tailoring product offerings and promotions.
5) Chatterjee and Das (2020) This research showed the application of clustering techniques on both demographic and behavioral data in order to differentiate between potential buyers and non-buyers in the vehicle sales industry.
6) Srinivas and Latha (2023) They applied Python-based machine learning algorithms, including K-Means, for customer segmentationindealershipdata.Theirworksupports the use of elbow methods

A. *Objectives of the study:*

1) To classify customers into different segmentsbasedondemographicfactors such as age, income level, and profession
2) To analyze the purchasing behavior of different customer segments and identify key patterns.
3) Toevaluatetheimpactofdemographics, financialstatus,andproductpreferences on customer segmentation.
4) Toapplymachinelearningtechniques (such as K-Means clustering) for automatedanddata-driven segmentation

B. Need for the study:

The increasing competition in the two-wheeler market demands a deeper understanding of customer behavior to retain and attract buyers. The market needs to take necessary steps to identifydistinctcustomergroupsandaddresstheir specific needs effectively. Customers exhibit diverse preferences and behaviors, which influence their decision to buy or not buy from a shop.Thegrowingcompetitionforcesmuchmore attention on understanding these differences to improve sales strategies. Customer segmentation helpsthebusinessbyenablingtargetedmarketing and optimizing resource allocation for future growth. This study, basedonthesegmentationof buyersandnon-buyersusingK-meansclustering, isquiteessentialfortheTVSSaihemanthmotors to thrive in the present-day competitive market.

C. Scope of the study:

The increasing competition in the two-wheeler market demands a deeper understanding of customer behavior to retain and attract buyers. The market needs to take necessary steps to identifydistinctcustomergroupsandaddresstheir specific needs effectively. Customers exhibit diverse preferences and behaviors, which influence their decision to buy or not buy from a shop.Thegrowingcompetitionforcesmuchmore attention on understanding these differences to improve sales strategies. Customer segmentation helpsthebusinessbyenablingtargetedmarketing and optimizing resource allocation for future growth. This study, basedonthesegmentationof buyersandnon-buyersusingK-meansclustering, isquiteessentialfortheTVSSaihemanthmotors to thrive in the present-day competitive market.

III. RESEARCH METHODOLOGY

A. ResearchDesign:

Exploratory Research

B. Data Collection:

Data were collected from 100 customers (50 buyers and 50 non-buyers) at TVS Sai Hemanth Motors in Srikalahasthi between March 30 andApril 12, 2025. Primary data were gathered through surveys, interviews, and feedback forms, capturing demographic details (age, gender, income, occupation, marital status) and preferences (style, speed, mileage, price, features, EMI, location, referrals, test rides). Secondary data were sourced fromthedealership'ssalesrecords,CRMsystem,and historical transaction data spanning the past 2-3 years. A convenience sampling technique was used dueto the accessibility and willingness ofcustomers to participate.

C. DataAnalysisTools:

Followingarethedifferenttoolsusedforanalyzing thedata
Dataanalyticaltoolsforclustering

- python
- Matplotliblibrary(usedforscatter plots, line graphs)

D. LimitationsofThe Study:

1) Thestudyisbasedononly50buyersand 50 non-buyers, which may not fullyrepresentalltypesofcustomersvisiting the dealership.
2) The segmentation is based only on demographic data. Behavioural, transactional,orservice-relateddatawere notincluded,whichcouldprovidedeeper insights.
3) Thestudyusesstaticdatacollectedatone point in time. Customer behaviour andpreferencesmaychangeovertime.
4) The findings are specific to one TVS Sai Hemanth motors located in srikalahasthi and may not be applicable to other regions with different customer demographics

5) The study doesn't include how buyers behave after the purchase (like service visits, satisfaction), which is crucial for long-term segmentation.

.

IV. DATA ANALYSIS & INTERPRETATION

A. DeterminingtheOptimalNumberofClusters–ElbowMethod

To perform effective customer segmentation using K-Means clustering, it is essential to determine the optimal number of clusters (K). The Elbow Method was used for this purpose. This technique involves plotting the Within-Cluster Sum of Squares (WCSS) against a range of K values

B. UseofPythonforEfficientClusteringandvisualization:

PerformingK-Meansclusteringandrelatedcomputationsmanuallyisnotonly
time-consumingbutalsopronetoerrors,especiallywhendealingwithmultiplevariablesand larger datasets. To overcome these limitations, Python programming was utilized in this
project.

C. BuyersDistribution Across Clusters:

After applying K-Means clustering, the customers were segmenteddistinct groups based on their demographic characteristics and preferences.Out of100 total respondents (50 buyers and 50 non-buyers), the distribution across

PYTHON CODE FOR CUSTOMER SEGMENTATION USING K-MEANS:

```
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.cluster import KMeans

# Encode categorical features
df_encoded = df.copy()
label_encoders = {}
for col in df.columns[1:]:
    le = LabelEncoder()
    df_encoded[col] = le.fit_transform(df[col])
    label_encoders[col] = le

# Standardize the data
scaler = StandardScaler()
X_scaled = scaler.fit_transform(df_encoded.drop(columns=['sl no']))

# Apply KMeans
kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
df_encoded['Cluster'] = kmeans.fit_predict(X_scaled)

# Analyze clusters
cluster_summary = df_encoded.groupby('Cluster').mean().round(2)
print(cluster_summary)
```

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Table1.Clustersolutionsforthebuyer'sdata

Feature	Cluster0	Cluster1	Cluster2
Age	1.0(18-25	2.5(26-35)	3.8(45+)
Gender	0.6	0.55	0.65
Qualification	1.2	2.1	3
Occupation	1.5	2.3	1
MaritalStatus	0.3	0.8	4
Annual Income	1.2	2.8	0.6
Style	0.5	0.7	0.4
Speed	0.3	0.6	0.5
Mileage	0.8	0.5	0.5
Price	0.6	0.7	0.7
Features	0.4	0.6	0.3
EMI	0.9	0.6	0.6
Location	0.7	0.5	0
Referral	0.82	0.6	0.4
TestRide	0.7	0.6	0.4

Table 2.InterpretationofFinalClusterCenters

Variable	Cluster0	Cluster1	Cluster2
Age	18-25Age group	Mostly26-35	mostly45+
Gender	60%aremale	Balancedgender	65%aremale
Qualification	MostlyUG students	MixofUG/PG job holders	Predominantly PG/PhD
Occupation	Students,entry- level job holders	Jobholders,self-employed	Business owners,senior professionals
MaritalStatus	Mostly unmarried (70%)	Mostlymarried (80%)	100%married
AnnualIncome	₹0–2.5L	₹2.5–10L	₹10L+
Style	Moderate interest(50%)	Highinterest (70%)	Stronginterest (83%)
Speed	Lowpriority (30%)	Highpriority (65%)	Moderate priority(40%)
Mileage	Critical(89%)	Moderate importance (50%)	Lowpriority (50%)
Price	Veryhigh(94% EMI users)	Moderate(60% EMI users)	Low(25%EMI users)

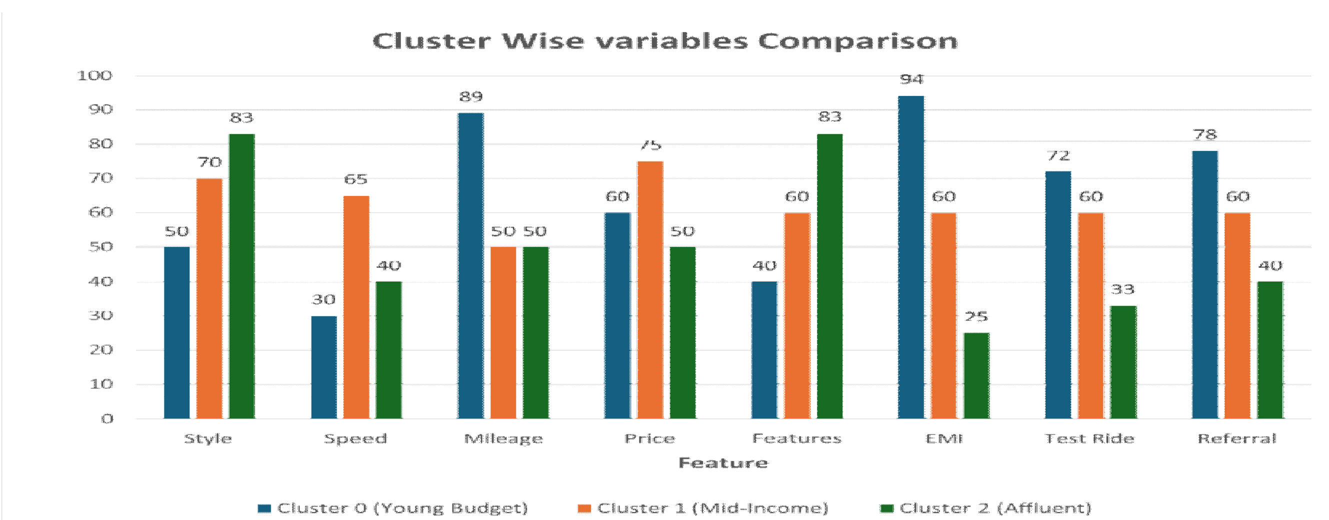
International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Features	Someinterest in features	Wantmore features	Lowpriority (30%)
EMI	StrongEMI preference	EMIinterestis moderate	EMIinterestis low
Location	61%Local, 39%Non-local	60%Local,40% Non-local	58%Local, 42%Non-local
Referral	Highly influencedby referral	Moderate(60%)	Low(40%)
TestRide	High(72%)	Moderate(60%)	Low(33%)

- Cluster 0: Representing young price-conscious buyers (36%), shows a strong preference for mileage (89%)andEMI(94%),reflectingtheirfocusonaffordabilityandfuelefficiencyasunmarriedstudentsor entry-level job holders aged 18-25 with low income (₹0-2.5L). This cluster also values referrals (78%), indicating trust in peer recommendations, but places less emphasis on style (50%) and speed (30%)
- Cluster1:Representingthemid-incomebalancedbuyers(40%),aged26-35withmoderateincome(₹2.5- 10L), demonstrates a more balanced preference profile, prioritizing price (75%), style (70%), and speed (65%), while showing moderate reliance on EMI (60%) and test rides (60%), suggesting a practical yet aspirational approach to purchasing.
- Cluster 2: Representing the affluent feature-seeking buyers (24%), aged 36-45+ with high income (₹10L+), exhibitsaclearpreferenceforstyle(83%)andfeatures(83%),withminimalrelianceonEMI(25%)ortestrides (33%), highlighting their focus on premium offerings and brand image over financing options

Table3.ClusteredDataSet

Numberofbuyersineachcluster			
CLUSTER	NOOFBUYERS	PERCENTAGE	igningbuyersof to clusters
0	18	36%	1,2,3,13,17,19,25,25,33, 40,41,48,50
1	20	40%	4,5,7,8,9,10,11,12,14,15, 16,20,21,22,24,26,29,32, 35,44
2	12	24%	6,18,23,27,30,31,34,37,38 38,42,43,47



Interpretation: The bar graph illustrating cluster-wise preferences at TVS Sai Hemanth Motors delineates distinct behavioral patterns across three customer segments for both buyers and non-buyers, offering valuable insightsfortargetedmarketing.Amongbuyers,Cluster0,youngprice-consciousbuyers(36%),aged18-25with incomes of ₹0-2.5L, prioritizes mileage (89%) and EMI (94%), reflecting their focus on fuel efficiency and affordable financing, while also valuing referrals (78%) but showing limited interest in style (50%) and speed (30%). Cluster 1, mid-income balanced buyers (40%), aged 26-35 with incomes of ₹2.5-10L, demonstrates balancedpreferencesforprice(75%),style(70%),andspeed(65%),withmoderateengagementinEMI(60%) and test rides (60%), indicating a pragmatic yet aspirational purchasing approach. Cluster 2, affluent feature- seekingbuyers(24%),aged36-45+withincomesexceeding₹10L,emphasizesstyle(83%)andfeatures(83%), with minimal relianceon EMI(25%)ortest rides (33%),highlighting apreference forpremium offerings. For non-buyers,Cluster0,youngprice-consciousnon-buyers(32%),aged18-25withincomesof₹0-2.5L,mirrors its buyercounterpartby prioritizing mileage(80%),EMI,and referrals (85%),with strong interest in test rides (70%), suggesting potential for conversion with enhanced trust-building measures. Cluster 1, mid-income uncertain non-buyers (44%), aged 26-35 with incomes of ₹2.5-6L, shows moderate interest in style (59%), speed (52%), and features (51%), but lower engagement in test rides (45%) and referrals (50%), indicating uncertainty in the buying process. Lastly, Cluster 2, affluent indifferent non-buyers (24%), aged 36-45+ with incomes of ₹6L-10L+, prioritizes style (72%) but exhibits low engagement with features (25%), test rides (20%), and referrals (30%), reflecting disinterest or unmet expectations. This graph underscores the varying engagement levels across segments, informing tailored strategies to address customer needs effectively.

D. Non-BuyersDistributionAcrossClusters:

After applying K-Means clustering, the customers were segmenteddistinct groups based on theirdemographic characteristics and preferences.Out of100totalrespondents (50buyers and 50 non- buyers), the distribution across

Table4.Clustersolutionsforthenon-buyers

Feature	Cluster0	Cluster1	Cluster2
Age	0.25	1.8	2.75
Gender	0.9	0.85	1
Qualification	1.2	1.75	2.5
Occupation	0.6	2.1	3
MaritalSta	0.3	0.9	1
AnnualInc	0.45	1.6	2.8
Style	0.4	0.7	0.85
Speed	0.3	0.65	0.45
Mileage	0.8	0.5	0.4
Price	0.75	0.6	0.25
Features	0.35	55	0.2
Location	0.6	0.55	0.4
Referral	0.85	0.5	0.3
TestRide	0.7	0.45	0.25

Table2.InterpretationofFinalClusterCenters

Variable	Cluster0–Budget-Conscious	Cluster1–Mid-Tier	Cluster2–Affluent
Age	18–25Agegroup	Mostly26–35	Mostly45+
Gender	90%aremale	85%aremale	100%male
Qualification	MostlyUGstudents	MixofUG/PGjobholders	PredominantlyPG/PhD
Occupation	Students,entry-leveljobholders	Self-employed,businessprofessionals	Businessowners,seniorprofessionals
MaritalStatus	Mostlyunmarried(70%)	Mostlymarried(90%)	Allmarried
AnnualIncome	₹0–2.5L	₹2.5–6L	₹6L–10L+
Style	Lowtomoderateinterest(40%)	Moderateinterest(70%)	Highinterest(85%)
Speed	Lowpriority(30%)	Moderatepriority(65%)	Lowtomoderate(45%)
Mileage	Criticalfactor(80%)	Moderateimportance(50%)	Lowpriority(40%)
Price	Veryprice-sensitive(75%)	Moderatepriceconcern(60%)	Lowsensitivity(25%)
Features	Someinterest(35%)	Moderateinterest(55%)	Lowinterest(20%)
Location	Mostlylocalcustomers(60%)	Local(55%)	40%arenon-local
Referral	Highinfluenceofreferrals(85%)	Moderateinfluence(50%)	Lowinfluence(30%)
TestRide	Highinterestintestrides(70%)	Moderateinterest(45%)	Lowinterest(25%)
EMI	StrongEMIpreference	EMIinterestismoderate	EMIinterestislow

- Cluster 0: Representing young price-conscious non-buyers, constitutes 32% of the non-buyer base, reflecting a significant group of younger individuals aged 18-25 with low incomes (₹0-2.5L), who prioritize affordability factors like mileage (80%) and EMI, yetremainunconverted,possiblyduetotrustor pricingbarriers.
- Cluster 1: Representing mid-income uncertain non-buyers, emerges as the largest segment at 44%, indicating that nearly half of the non-buyer population consists of individuals aged 26-35 with moderate incomes (₹2.5-6L), who exhibit uncertainty in the buying process, as evidenced by their moderate engagement with style (59%), speed (52%), and test rides (45%).
- Cluster2: Representing affluent indifferent non-buyers, accounts for the smallest share at 24%, mirroring the proportion of affluent buyers, but these high-income individuals aged 36-45+ (₹6L-10L+) show disinterest through low engagement with features (25%) and test rides (20%), suggesting unmet expectations or lack of motivation to purchase.

Table3.ClusteredDataSet

Cluster	NonBuyerCount	Percentage	Assigned non buyerstoclusters
0	16	32%	1,3,6,10,13,15,16,17,19,24,25,28,33,39,41,45

1	22	44%	2,4,5,7,8,9,11,12,14,20,21,22,23,26,29,30,31,32,34,35,40,46
2	12	24%	18,27,36,37,38,42,43,44,47,48,49,50

V. FINDINGS

TheElbowMethodconfirmedthatK=3is optimal for both buyers and non-buyers, enabling meaningful segmentation into three distinct customer types in each group

1) Buyer Segmentation:
- Cluster0:Young,low-incomeindividuals focused on price and mileage.
- Cluster1:Mid-incomeprofessionalswho balance cost, style, and EMI options.
- Cluster 2: High-income buyers highly influencedbyfeatures,style,andtestrides.

2) Nonbuyer Segmentation:
- Cluster0:Studentsoryoungindividuals exploring options with low intent.
- Cluster1:Moderatelyinterestedindividuals who might buy with the right offers.
- Cluster2:Uninterestedgroup,mainlyolder individuals with low engagement.

3) Comparision:
- Buyers are more engaged with features, Emi & test rides
- Non buyers lacked motivation due to income limitations and awareness

VI. SUGGESTIONS

1) Offer Deals to Cluster 0 Non-Buyers: These people care about EMI and mileage. Give them low EMI and fuel-efficient options to encourage buying.
2) Support Cluster 1 Non-Buyers: They're unsure. Let them try test rides and explain features clearly to build confidence.
3) Attract Cluster 2 Non-Buyers: They have money but low interest. Show premium features or special deals to catch their attention.
4) UseReferrals&TestRides:Encourageword- of-mouth and offer test rides to increase customer interest.
5) Send Right Offers to Right Groups: Use cluster data to match offers with the right people for better results.

VII. CONCLUSION

K-Means clustering algorithm is successfully appliedto segment bothbuyers and non-buyers based on demographic and behavioral variables. The analysis revealed clear patterns in customer preferences and highlighted key differencesbetween similar customer groups who did and did not make a purchase. These insights help the dealers understand their customers better and craft targeted strategies for each segment. By bridging the gap between buyers and non-buyers through focused offers and engagement, the dealership can increase conversions and optimize its marketing efforts. Overall, this study demonstrates the practical value of machine learning in solving real- world business problems

REFERENCES

[1] BagirathiIyer,(June2011),AStudyonReasons for Low Acceptability of Battery Operated 2 Wheelers (E-Bykes) in Nagpur Region with Special ReferencetoElectrothermIndiaLimited's Yobykes, International Journal of Research in Finance & Marketing, Volume1, Issue2.
[2] PhilipKotler,(2003),MarketingManagement, 11thEdition, Pearson Eduction, Delhi, India, pp:278- 299.
[3] NareshK.MalhotraandSatyabhushanDash,(2007), MarketingResearch-AnAppliedOrientation, Fifth Edition, Pearson Education, New Delhi, India, pp. 636-661.
[4] M.N. Tuma, S. Scholz, and R. Decker, (2009), The Application of Cluster Analysis in Marketing Research: A Literature Analysis, Business Quest Journal, vol. 14.

[5] M MKodabagi, Savita S Hanji, Ravatappa A.B,(2013) "This is my paper", A Novel Clustering Technique Using Similarity Measure, International Journal of Latest Trends in Engineering and Technology (IJLTET), Special Issue IDEAS-2013,ISSN: 2278-621X.

[6] H. N. Ramesh, Vinod S., Sanjay Hanji, (Jan-June 2012) "This is my paper", Hybrid Segmentation AtTwo Wheeler Market In India, IEMS Jounal of Management Research, ISSN : 2249-569X,Volume1, Issue-1.

[7] RakeshKumarArora,Dr.DharmendraBadal,(April- June 2013), Evaluating Student's Performance Using k-Means Clustering, International Journal of Computer Science AndTechnology(IJCST), ISSN(ONLINE):0976-8491ISSN(PRINT):2229-4333Vol.4,Issue-2.

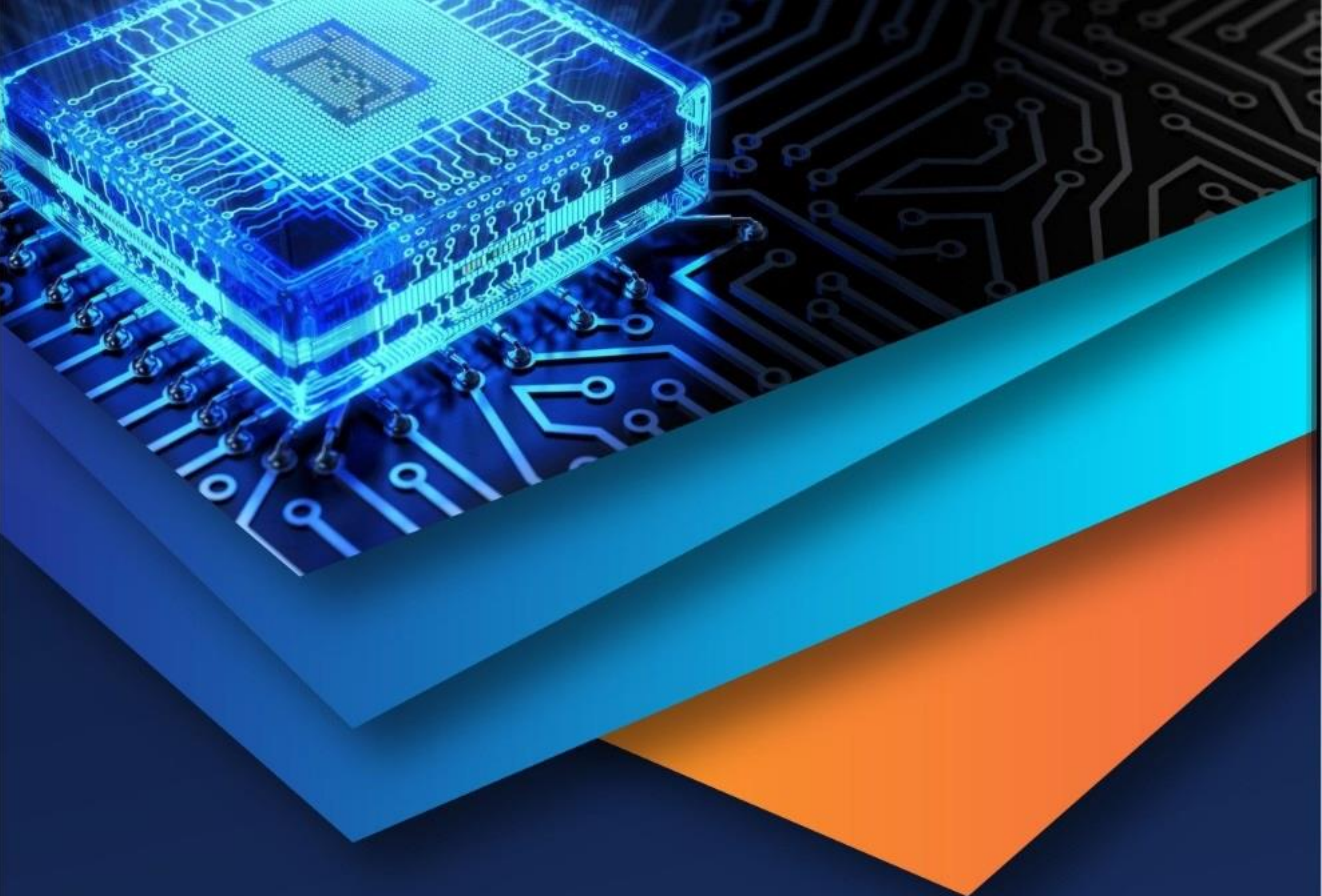
[8] Hedayetul Islam Shovon, Mahfuza Haque, (2012), An Approach of Improving Student's Academic Performance by using K-means clusteringalgorithm and Decision tree, International Journalof Advanced Computer Science and Applications(IJACSA), Vol. 3, No. 8.

[9] Chady El Moucary, Marie Khair, Walid Zakhem, (2011), Improving Student's Performance Using Data Clustering and Neural Networks in Foreign- Language Based Higher Education, The Research Bulletin of Jordan ACM, Vol. I (III).

[10] Oyelade, O. J, Oladipupo, O. O, Obagbuwa, I. C, (2010), Application of k-Means Clusteringalgorithm for prediction of Students' Academic Performance, International Journal of Computer Science and Information Security (IJCSIS), Vol. 7.

[11] Md. Hedayetul Islam Shovon,MahfuzaHaque,(July 2012), Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm, International Journal of Advanced Research in Computer Science and SoftwareEngineering,ISSN:2277128XVol2, Issue 7.

[12] Mahmud M.S, Rahman M.M, Akhtar M.N, (2012), ImprovementofK-meansclusteringalgorithm with better initial centroids based on weighted average, Electrical & Computer Engineering(ICECE), 7th International Conference, Pages:647- 650.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY