iJRASET

International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: VII Month of publication: July 2021

DOI: https://doi.org/10.22214/ijraset.2021.37134

Behavioral Based Credit Card Fraud Detection

Shalini S[1], Deepika G. D[2], Shalini M.R[3], Nivedidha R[4], N Bhavani[5]
[1, 2, 3, 4, 5]Saranathan College of Engineering, India

Abstract: Credit card fraud is a significant threat in the BFSI sector. This credit card fraud detection system analyzes user behavioral patterns and their location to identify any unusual patterns. This consists of user characteristics, which includes user spending styles as well as standard user geographic places to verify his identity. One of the user behavior patterns includes spending habits, usage patterns, etc. This system deals with user credit card data for various characteristics, which includes user country, usual spending procedures. Based upon previous transactions information of that person, the system recognizes unusual patterns in the payment method. The fraud detection system contains the past transaction data of each user. Based on this data, it identifies the standard user behavior patterns for individual users, and any deviation from those normal user patterns becomes a trigger for the detection system. If it detects any unusual patterns, then user will be required to undergo the security verification, which identifies the original user using QR code recognition system. In case of any unusual activity, the system not only raises alerts but it will block the user after three invalid attempts.

Index Terms: Supervised Learning Algorithm, credit card, fraud detection, fraudulent prediction, Security Verification.

I. INTRODUCTION

The usage of credit cards has dramatically increased nowadays. A credit card is a thin plastic card that contains information such as a signature or photo and it authorizes the person named on it to charge purchases or services to his account and imposes for it, which he will be billed periodically. Now the information on the card is read by ATMs, store readers, banks and is used in an online internet banking system. It contains a unique Card number, which is of utmost importance. The security relies on the physical security of the plastic card and the privacy of the credit card number. Due to the rise of electronic commerce, there has been a tremendous use of credit cards has been increased that led to a huge amount of frauds related to the credit cards. Credit card fraud happens when someone uses our credit card or credit account to make a purchase that we did not authorize.

It may happen in many ways:

1) If we lose our credit card or have it stolen, it can be used to make purchases or different transactions, either in person or online.
2) Scammers can also steal the credit card account number, PIN and security code to make unauthorized transactions, without needing your physical credit card. In the digital world, it is mandatory to detect credit card frauds. Fraud detection includes monitoring the behavior of various users and analyze it to estimate, detect or avoid undesirable behavior. The objective of our system is to identify the suspicious events and reduce the theft or fraud from being happened and to reduce the risky work of an employee who works in the bank.

II. LITERATURE SURVEY

In SPManiraj@et.al paper, it is vital that credit card companies are able to identify fraudulent credit card transactions so that customers are not charged for things that they did not purchase. Such issues is tackled with Data Science and its importance, along with Machine Learning, cannot be overstated. Their project intends may be illustrate the modelling of a data set using machine learning with Credit Card Fraud Detection. The drawbacks of Credit Card Fraud Detection include modelling the previous credit card transactions with the data of the ones that turned out to be fraud. This model is then used to find whether a new transaction is fraud or not. Their objective here is to find 100% of the fraudulent transactions whereas minimizing the wrong fraud classifications. The Credit Card Fraud Detection is a common example of classification. In this process, they have focused on analyzing and pre-processing data sets as well as the deployment of multiple anomaly detection algorithm includes Local Outlier Factor and Isolation Forest algorithm on the PCA transformed Credit Card Transaction data. [9]

In DVarmedja@et.al paper, the research shows several algorithms that can be used for classifying transactions as fraud or genuine one. Credit Card Fraud Detection dataset was employed in the research. Because of highly imbalanced data, SMOTE technique was used for oversampling. Further, feature selection was performed and the dataset was split into two parts, training data and test data. Logistic Regression, Random Forest, Naive Bayes and Multilayer Perceptron were the algorithms used in this experiment. It shows that each algorithm can be used to detect credit card fraud with high accuracy. The proposed model can be used to detect other irregularities. [12]

In VaishnaviNarthDornadula@et.al study, the main aim is to design and develop a novel fraud detection method for streaming Transaction Data, with an objective, to analyze the past transaction details of the customers and extract the behavioral patterns. Where cardholders are divided into different groups according to their transaction value. Then using a sliding window strategy to aggregate the transaction made by the cardholders from different groups so that the behavioral pattern of the groups can be extracted respectively. Later different classifiers are trained over the groups separately. And then a classifier with better rating points can be chosen as one of the best ways to predict fraud. Thus, followed by a feedback mechanism to resolve the problem of concept drift. In this paper, we worked with the European credit card fraud dataset. [11]. In ShiyangXuan@et.al paper, two types of random forests are used to train the behavioral characteristics of normal and abnormal transactions. They make a comparison of the two random forests, which are different in their base classifiers, and analyze their performance on credit fraud detection. The data used in their experiments come from an e-commerce company in China.[10]

III. RELATED STUDIES

A. Machine Learning

Machine learning is a subfield of Artificial intelligence that permits computers to take as input some data and makes decisions by learning from it. It is the scientific study of algorithms and static models that computer systems use in order to perform a specific task effectively without using the explicit instructions, relying on patterns. In traditional programming, we give the input and logic and machine runs to get the output. But in machine learning we give input and output, run the machine during training and machine itself creates an own value, that can be evaluated while testing Machine learning algorithms build a mathematical model based on sample data, known as training data in order to make predictions and decisions.

There are two basic types of machine learning,

1) *Supervised Learning:* In supervised learning the machine learns from the trained data we are given as input and output.
2) *Unsupervised Learning:* In this type of learning the machine foretell the output on its own by comparing the older values. It does not require the training for the process.

The algorithms used in machine learning are Linear Regression, logistic regression, decision tree, Naive Bayes, K-means, Random forest.

The main applications of machine learning are

a) Email filtering
b) Transactions
c) Speech and face recognition
d) Chabot
e) Self-driving cars
f) Fraud detection, etc.

B. Random Forest Algorithm

Random forest algorithm is a supervised learning algorithm, which comes with the advantage that it can be, used for both classification and regression problems. As the name suggests it builds multiple decision trees and merges them along to get more accurate and stable predictions. Random forest adds additional randomness to the model, while growing the tree instead of searching for the most important feature while a node searches for the best feature. In our model, we have four features like geographic location, amount, frequency of spending and vendor. The model builds a separate decision tree for each feature and combines them using an ensemble, which predicts an accurate result for the given transaction. Since the random forest combines multiple decision trees to predict class of the dataset, it is possible that some decision trees could predict the proper output, whereas others may not. However, all the trees predict the proper output.

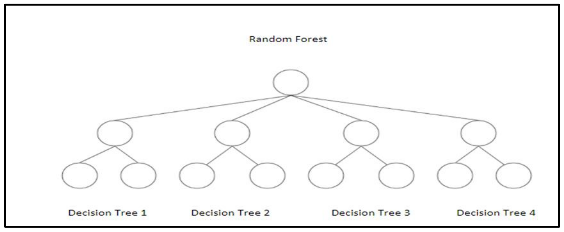


Fig 1: Random Forest Algorithm

C. Behavioral Pattern

The behavior pattern is one of the topics well studied in computer perspective. Behavior design patterns designed to be a constructive pattern in software engineering; this recognizes the pattern of normal fluctuations between objects and recognizes them. Because of this ever-increasing rise in social media. Authority between objects or behavior that binds to an object and gives it requests stresses patterns of behavior. In addition, they are concerned about their interactions, dependence on others, and ways to differentiate themselves independently and rely on providing flexibility. In these patterns, the interaction between objects should be such that they can work together easily and mean that they should still be freely integrated. Free integration should be mentioned to avoid coding hard and dependable. There are various types of patterns namely Chain of Responsibility, null object, Command, visitor, Interpreter, strategy, Iterator, Mediator, state, Memento, Observer and Template pattern.

IV. METHODOLOGY

Data is collected in the form of a dataset. This dataset is now split into part for training and testing in a particular ratio. The Random forest classification algorithm creates a classifier model. This model is trained in a training database and tests are performed using the test dataset. the training and testing model together predict and provide an accurate result as fraud or legal.

1) Collecting Data: Data is collected and assembled in the form of a CSV file, which contains the details of transactions made by European credit card holders in September 2013. The data are mentioned in table format which contain time, v1, v2, v3…v28, amount, class.
2) Model Split: We split the dataset at a rate of 67:33 (train: test) ratio, in which 67% of the data is given to model training and 33% of the data is used for model testing.
3) Random Forest Algorithm: A Fraud prediction model is implemented based on Random Forest algorithm and the remaining testing data is tested using the model.
4) Model Evaluation: At last, a confusion matrix is calculated based on the calculated value.

A. Dataset

The dataset that we use in our model contains transactions made by European credit card holders in September 2013. This dataset contains 284,807 transactions of which 492 were fraudulent and the remaining 284,315 were recorded as legitimate. As it is clear from the count that the data is highly imbalanced, fraudulent transactions account for only 0.17% of total transactions. Features from V1, V2, V3… V28 were detected by PCA (Principal Component Analysis). Attributes such as Time and class have not been changed using PCA. Time indicates the time (in seconds) that is taken between the transactions. Amount is the transaction amount that was made.

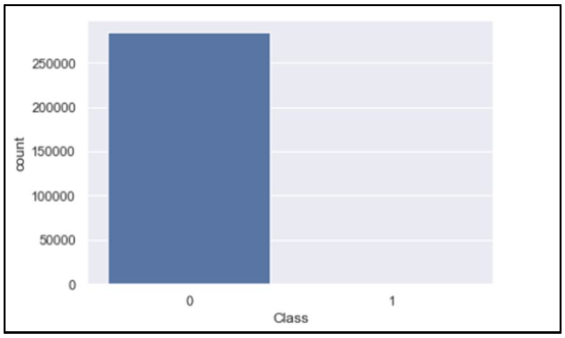


Fig 2: Dataset Analysis

The graph obtained above is unbalanced and uncertain because the dataset contains 284,807 transactions out of which only 284,315 are legitimate and 492 are fraud transactions. We have 0.999 score on accuracy, 0.937 score on precision, 0.798 score on recall and 0.862 score on F1 score. Hence, we perform under sampling techniques to balance those uncertain data to have an accurate prediction.

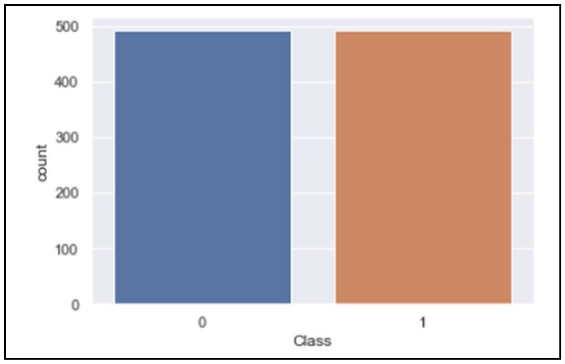


Fig 3: Dataset Analysis after under sampling

The dataset we had already was based on those, which already existed and had less minority class data in the dataset. After applying the under sampling technique, the majority class becomes equal to the minority and this randomly picks a point from the minority class for computing the fraudulent prediction.

B. Proposed Model

In the proposed system, analyzing location and behavior are used. Our system uses user behavior and location scanning to check the fraudulent transaction. This checks the users spending pattern and their location to verify their identity. If any suspicious event is detected, the system asks for security verification. The system verifies the user using QR code. If a fraud detection system identifies the fraud, then it will send an alert to the original user and blocks the user for more than three
Invalid attempt.

1) System Architecture

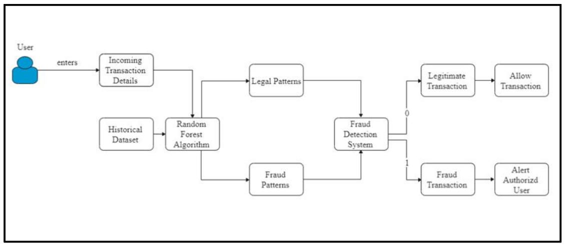


Fig 4. System Architecture Diagram

User enters the transaction details into the system. These transaction details and the dataset are the input to the Random Forest algorithm, which builds multiple decision trees and merges them along to get a more accurate and stable prediction. It will predict the legal patterns and fraud patterns. The fraud detection engine verifies the user behavior pattern. If any deviation occurs, then the user should undergo security verification. If the given details are correct, then it will allow users to proceed the transaction otherwise, it will block the user. An alert message will be sent to the original user, if a fraud transaction is detected.

V. MODULES

In general, there are four modules used in our proposed system namely,

A. User Module

In the user module an interactive front-end page using HTML and CSS is used which allows the customers to enter their details like user name and password, in case of new customers, they are allowed to register their candidature in the future. After every registration, customers are supposed to login again for security purposes. Customer details like their name, DOB, Account number, Mobile number, Branch code, Card Number, CVV code and Password are fetched. The details that are collected from the user are checked using JavaScript as a client-side validation and stored in a database, which makes it easier for retrieval and accessing the data.

B. Payment Module

Transactions to be made by the user are carried out in this module; the user is allowed to enter the receiver account details like receiver account number, name, IFSC code and amount to be transacted. On submitting, the details are verified at the background and if the given credentials are true, the amount is transferred, else users need to undergo some security verification process, which in terms gives the confirmation that it is a legal and authorized transaction.

C. Fraudulent Prediction

This module identifies the fraud transaction using Random Forest Algorithm and the datasets that are fed into the system.

1) *Fraud Detection:* We nearly have four features in the system. The system builds a separate decision tree for each feature and then merges by ensemble model to build a complete random forest algorithm. This is used to predict the incoming transaction.

a) *Geographic Location:* This feature is used to track the location where the transaction is being made and analyze the user's IP address.

b) *Amount:* This feature keep in track of the amount that is being transacted to the respective user

c) *Frequency of Spending:* This feature is used to analyses the spending pattern of a specific user.

d) *Vendor:* Which keeps in track of the receiver accounts.

2) *How Random Forest Works:* Random Forest works in two-phase: first is to create the random forest by combining N decision trees, and second is to make predictions for every tree created in the first phase. The working process of Random Forest starts by selecting random. Samples from the given dataset. Next, the Random Forest algorithm will construct a decision tree for Every sample. Then it will get the prediction result from each decision tree. Next voting I will be performed for every predicted result. At last, select the most voted prediction result as the final prediction result.

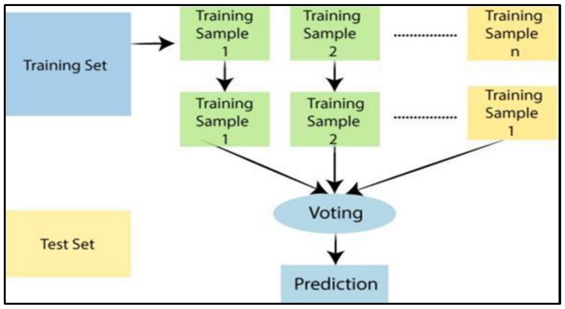


Fig 5. Random Forest Algorithm Working

D. Security Verification Module

In case of any deviation in processing the transaction is identified, users need to undergo some security verification process to proceed further. Verification is done by QR code recognition of an user who is making transaction, if the user who is attempting to make transaction deny to undergo verification or if any suspicious act has been monitored by the system it sends an alert message to the user and more than three invalid attempts blocks the particular user from making a transaction. We use OpenCV for detecting the QR code. The camera locates and captures the QR code of the user. The frame is passed to a dedicated Python barcode decoding library such as a Zbar. The ZBar library will then decode the QR code. OpenCV is used for further processing and for displaying the result. If the QR code of the user matches an image in a database, then the user is identified as a legitimate user.

VI. CONCLUSION

We proposed Credit Card fraud detection for detecting frauds based on their behavior pattern. It has been demonstrated that our models are able to detect fraud based upon their spending pattern etc. The fraud detection system can easily evaluate the fraud based on the Random Forest Algorithm, and our project will verify the user as fraud or not by the user's QR code.

VII. FUTURE ENHANCEMENT

The project will be very much useful in this challenging real world. More functionality can be added in accordance with the flexibility of the user requirement and specification. It has a vast scope in future. In future, this project can be improved by using face recognition for verification purposes.

REFERENCES

[1] ”Transaction-Level Behavior Based Credit Card Fraud Detection Mechanism by Bhakti Ratnaparkhi, Rahul Patil” by (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6), 2014, 7071-7074

[2] QR code recognition based on principal components analysis method by HichamTribak, Youssef Zaz Int. J. Adv. Comput. Sci. Appl. (IJACSA) 8 (4), 241-248, 2017

[3] Credit card fraud detection-by ishutrivedi, Monika, mrigya, mridushi” published By International journal of advanced research in Computer and communication Engineering

[4] Credit card fraud detection based on transaction behavior-by john Richard D. Kho, Larry A. Vea,” published by proc. Of the 2017 IEEE region 10 conference (TENCON), Malaysia, November-5-8, 201

[5] “Analysis of Spending Pattern on Credit Card Fraud Detection by Capt. Dr. S Santhosh Baboo, N Preetha “by IOSR Journal of Computer Engineering (IOSRJCE) 2015

[6] ”Detecting credit card fraud by decision trees and support vector machines by Yusuf G Şahin, EkremDuman” by Newswood Limited, 2011

[7] REAL-TIME CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING by Anuruddha Tennakoon, Chee Bhagyani, SasithaPremadasa, ShalithaMihiranga, NuwanKuruwitaarachchi by IEEE 2018

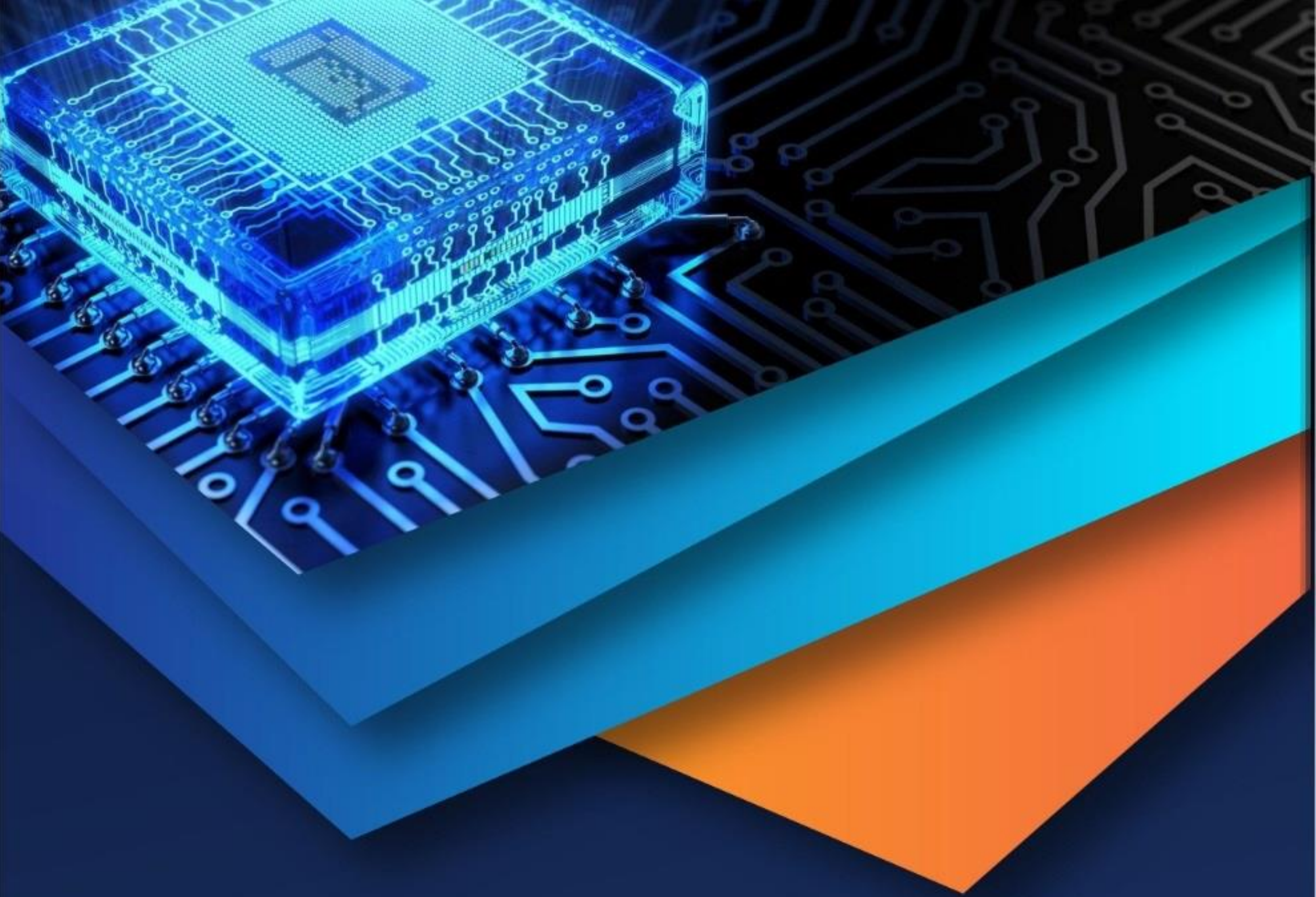
[8] CREDIT CARD FRAUD DETECTION USING ADABOOST and MAJORITY VOTING by KULDEEP RANDHAWA 1, CHU KIONG LOO 1, (Senior Member, IEEE), MANJEEVAN SEERA 2, 3, (Senior Member, IEEE), CHEE PENG LIM4, AND ASOKE K. NANDI5, 6, (Fellow, IEEE) by IEEE Access 2018

[9] CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING AND DATA SCIENCE by S P Maniraj , Aditya Saini , Shadab Ahmed , Swarna Deep Sarkar by IEEE 2019

[10] RANDOM FOREST FOR CREDIT CARD FRAUD DETECTION by Shiyang Xuan , Guanjun Liu , Zhenchuan Li , Lutao Zheng , Shuo Wang , Changjun Jiang By IEEE 2018

[11] CREDIT CARD FRAUD DETECTION – MACHINE LEARNING by VaishnaviNarthDornadula, Geetha S by IEEE2019

[12] CREDIT CARD FRAUD DETECTION – MACHINE LEARNING METHODS by D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic and A. AnderlaBy IEEE 2019.



doi crossref

10.22214/IJRASET

INDEX COPERNICUS

45.98

ISRA JIF

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY