iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 9 **Issue:** VI **Month of publication:** June 2021

**DOI:** https://doi.org/10.22214/ijraset.2021.35377

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429*
*Volume 9 Issue VI Jun 2021- Available at www.ijraset.com*

# Sentiment Analysis of College Reviews using Machine Learning

Toshita Chandurkar[1], Dr. Pritish Tijare[2]
[1]*ME Student,* [2]*Associate Professor, Department of Computer Science and Engineering, Sipna College of Engineering and Technology, Amravati, Maharashtra, India*

***Abstract:*** ***Sentiment analysis is the process of detecting positive or negative or neutral sentiment in text. It's often used by businesses to detect sentiment in social data, gauge brand reputation, to understand it and make better. Sentiment analysis models focused on polarity (positive, negative and neutral) and even intentions (interested or not interested). Depending on how we wish want to interpret feedback and queries, we will define and tailor your categories to meet your sentiment analysis needs. This paper focuses the reviews of various colleges which are an important form of opinion mining. The essential objective of this paper is to classify every sentence's semantic orientation (e.g. positive, negative and neutral) of the reviews. It is a very useful analysis since we could possibly determine opinion about various colleges.***
***Keywords: Sentiment Analysis, Machine Learning, Opinion Mining.***

## I. INTRODUCTION

Nowadays the undeniable truth is that social media users are more likely to share how they feel about a couple of current "hot topic" on social media platforms. Users supposed to post positive, negative or neutral opinions about that topic or as specific product they are using. The advancement of Machine Learning has opened doors in detect and categorize online opinions [1]. We analyze the opinion of user views about the recently controversial issues and compare them with the related different major issues like campus placement, infrastructure facilities, LAB facilities, Teaching methodology, etc. Analysis of sentiments is basically the method of deciding whether the sentiment in the text is positive, negative or neutral.

Existing research has produced numerous techniques for various tasks of sentiment analysis, which include both supervised and unsupervised methods [2]. In the supervised setting some papers used all types of supervised machine learning methods (such as Support Vector Machines (SVM), Maximum Entropy, Naïve Bayes, etc.) and feature combinations. About a decade ago, deep learning has emerged as a very powerful machine learning technique and produced state-of-the-art results in specific domains, valuing from computer vision and speech recognition to NLP. Machine learning approach is proposed for the efficient classification of college data. To increase the efficiency of classification process, the specific processing technique is implemented at each node. After training the dataset the machine is analyses with various set of test data [3].

## II. LITERATURE REVIEW

Numerous studies have been done that focus on Sentiment Analysis which is a research branch of text mining; they have applied on different probabilities for different methods. Nageswara Rao Moparthi et al. [2] had shown Movies Reviews Sentiment Analysis and Classification in which they perform a task Different classification algorithms are considered and compared to assess their performances for the task at hand. The reason why NB, BN, KNN, SVM and SGD classifiers were chosen to be compared with each other is that these algorithms are supervised classifiers that have proven their efficiency and reliability in SA based on the previous works studied. Wei Zhao, et al. [3], Weakly-Supervised Deep Embedding for Product Review Sentiment Analysis reflects the general sentiment distribution of sentences, from a large number of weakly labeled sentences.

## III. PROBLEM ANALYSIS

Today Sentiment Analysis is widely used in many sectors, but the present system is getting old and new generation faced many problems and also the main problems that exist within the current techniques are: inability to perform well in different domains and give inadequate accuracy. Performance in sentiment analysis based on insufficient labeled data, incapability to deal with complex sentences that require more than sentiment words and easy analyzing, but in some cases there is problem while considering negative sentences [3]. So there is scope to implement a new system within the direction of reducing computational complexity and accurate output.

## IV. SENTIMENT ANALYSIS

Sentiment analysis of various college review technique is one of the prediction detection systems. The advantage of this is to know public opinions and extract their emotions by considering them and explained how database gives advantage during admission process. This method suggested that the way to choose good one for bright future. It opted two stage approach for their framework- first preparing training dataset from given database using machine learning relevant features, after collecting and preprocessing the dataset training data set was created [4]. Next by using Sentimental Analyzer output gets the polarity in percentage. This approach reduced the number of training set and further they applied Machine learning approach and classification algorithm to get the polarity of database.

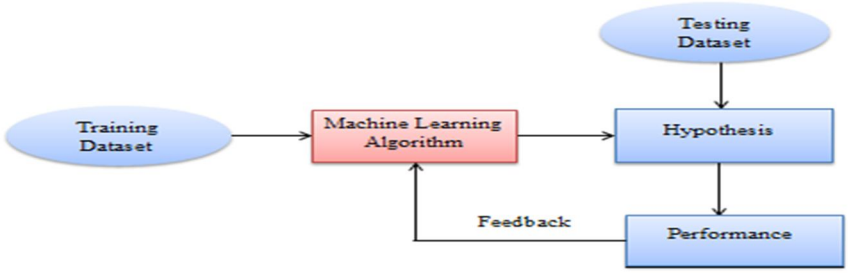



Figure 1: Sentiment Analysis Using Machine Learning

The common approach found in almost most relevant research works constitutes data collection, preprocessing and filtering of data then approaches in feature extraction, classification and pattern analysis makes the distinction [5]. This kind of information forms a basis for people to judge, rate about the performance of not only any college but about other products and to know about whether it will be a successfully or not. Therefore, sentiment analysis has wide applications and includes emotion mining, polarity and influence analysis [6].

## V. PROPOSED WORK

In this paper for the analysis and classification of various college data with various attributes is considered. This dataset contain 8 attributes we refer all of them to generate result. This dataset is freely available to the internet. A comment contains a lot of opinions about the data which are expressed in different ways by different users. The dataset used in this survey work is already labeled into classes' viz. negative, positive and neutral polarity and thus the sentiment analysis of the data becomes easy to observe the effect of various features.

The workflow of project by using Sentiment Analysis given in figure 2, which provide the brief overview of fundamental steps that should be followed to apply machine learning algorithm on dataset:

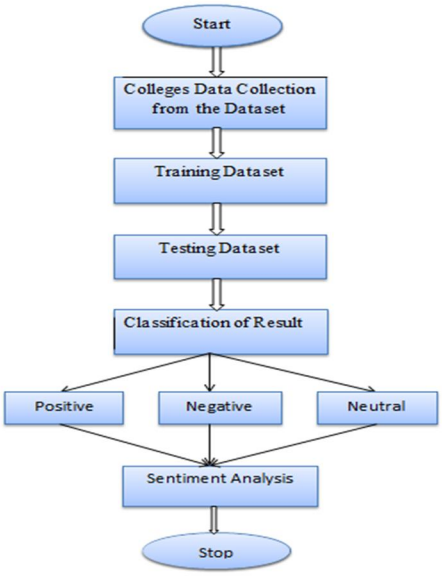



Figure 2: Fundamental Steps for Training and Testing

*A. Process Of Sentiment Analysis*

1) *Data Collection:* Dataset is in the form of raw comments which is retrieved. This dataset is nothing but the reviews on colleges stored.
2) *Training Dataset:* Data in the training dataset must be pre-processed before the evaluation by the Sentiment analysis. It involves Tokenization which is the process of splitting the comments into individual words called tokens.
3) *Testing Database:* Test data is used to measure the accuracy and efficiency of the algorithm used to train the machine - to see how well it will predict new answers based on its training method.

The processing of machine learning technique to generate Sentiment Analysis,

*a) Raw Dataset:-* Initially select complete Raw Dataset which contain all 8 attribute with punctuation and interrupted sign.
*b) Pre-processing Module:-* Collect Raw Dataset remove all interrupted sign (*, &, %, $, #, @, !, etc.), and generate comma separated value (CSV) file. With this generate sentiment (1, -1, 0) according to comment.
*c) Processed Dataset:-* After Pre-processing Processed Dataset generated which will divides into Testing and Training dataset.
*d) Train/Test Dataset:-* In this dataset will be test and train in three categories i.e. 80%-20, but it is random selection of dataset in both Random Selection case and College Selection case.
*e) Random/College Selection Result:-* Training and Testing process generate the accuracy of model which is new up to first 5 cases and next test accuracy will be reflected any value from above.
*f) Analyze College Data:-* It shows drop-down list of all colleges, after selecting any one college gives us all attribute wise sentiment analysis and justify which college is best in which field.

Based on the data of confusion matrix, precision, recall, F-measure and accuracy are the evaluation measures used for evaluating performance of classifier.

- *Precision:* It measures the exact value of the classifier result. For binary classification problem, precision is the ratio of number of reviews correctly labeled as positive.

$$\text{Precision} = \frac{TP}{TP+FP} \quad [7]$$

- *Recall:* It measures the complete value of the classifier result.

$$\text{Recall} = \frac{TP}{TP+FN} \qquad \text{TNR} = \frac{TN}{TN+FP} \quad [7]$$

Where,
TP - True Positives,
TN - True Negatives,
FP - False Positive and
FN - False Negatives

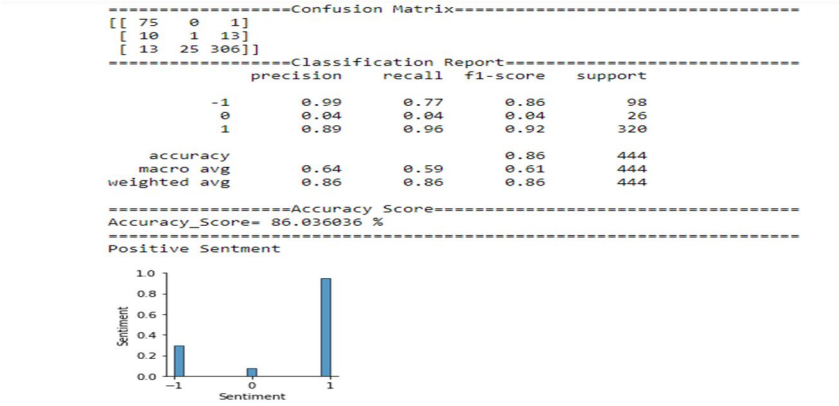
- *F-measure:* It is the mean value of precision and recall.

$$\text{F - Measure} = \frac{2 * Precision * Recall}{Precision + Recall} \quad [7]$$

- *Accuracy:* It can be calculated as the ratio of correctly classified reviews to total number of reviews.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad [7]$$

- *Confusion Matrix:* It is used for evaluating the performance of the classification model.

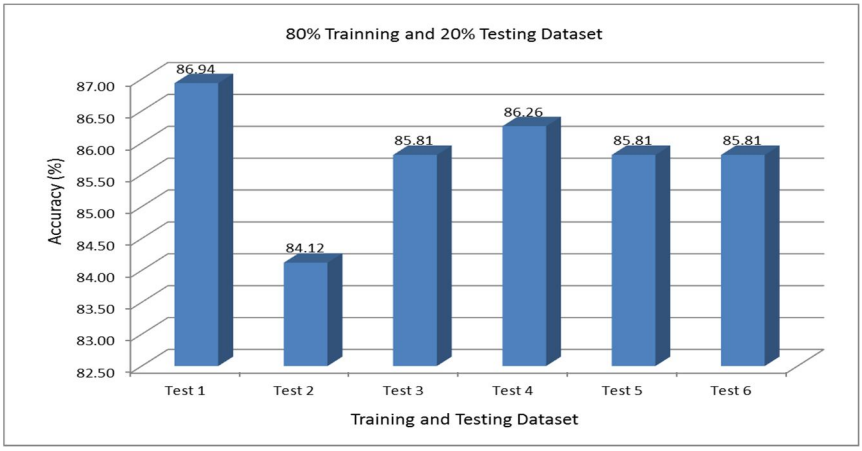



Screenshot 1: Application of training and testing

## VI. RESULTS

### A. *Experiment for Random Selection of Dataset*

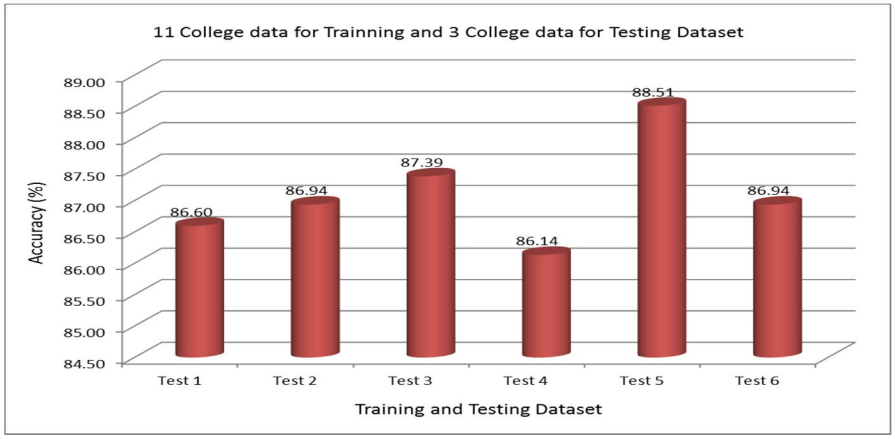
Classification of Tangle feedback dataset for training (80%) and testing(20%) using Machine Learning Program (MLP), the result is as follows,




Graph 1: Percentage Accuracy for Random selection of training (80%) and testing (20%)

### B. *Experiments for College wise Selection of Dataset*

Classification of Tangle feedback dataset for training (80%) and testing (20%) using Machine Learning Program (MLP), the result is as follows,




Graph 2: Percentage Accuracy for College wise selection of training (80%) and testing (20%)

## VII. APPLICATION

Sentiment analysis is applicable to many real world business problems and marketing such as social media monitoring, twitter comment, customer feedback etc. [8]. In the background of sentiment analysis, advanced AI algorithms apply language deconstruction techniques, like tokenization, part-of-speech tagging, parsing, and lemmatization to break down and make sense of text [9]. Only the machine learning software classifies unstructured text by emotion and opinion mining. Sentiment Analysis is a term that includes various tasks such as sentiment extraction, sentiment classification and summarization of opinions or opinion spam detection, sentiment generation among others. It aims to analyze people's attitudes, opinion and emotions, etc. based on peoples comments [10].

## VIII. CONCLUSION

Sentiment Analysis has led to development of better products, public behavior and good business management reviews. In this paper, we tried to show the basic way of classifying comments into positive, negative, neutral category using Machine Learning as baseline. We have performed conclusion examination for utilizing colleges' information [11]. Therefore, look at the details, make a selection of the data, apply transformations and filter the more relevant data making machine learning methods generalize and effective since the computers these days have limits and can't handle them all the data without prior review of any kind [12]. It can find unfair positive reviews or unfair negative reviews or unfair neutral reviews, reputation issues, and collusion and control through this work.

## REFERENCES

[1] Toshita Chandurkar, Dr. Pritish Tijare, "Sentiment Analysis: A Review and Comparative Analysis on Colleges", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), UGC Approved – 64718, ISSN: 2456-3307, Volume 7 Issue 2, pp. 526-531, March-April 2021.

[2] Nageswara Rao Moparthi, Dr. N.Geethanjali, "Design and implementation of hybrid phase based ensemble technique for defect discovery using SDLC software metrics", An International Conference by IEEE, AEEICB16 (978-1-4673-9745-2) PP. 269-276, IEEE 2018.

[3] Wei Zhao, et.al., "Weakly-Supervised Deep Embedding for Product Review Sentiment Analysis", IEEE Transactions on Knowledge and Data Engineering, Vol. 30, No. 1, 185, 2018.

[4] Rui Xia, Feng X, ChengqingZong, Qianmu Li, Yong Qi, Tao Li,Dual, "Sentiment Analysis: Considering Two Sides of One Review", IEEE Transactions On Knowledge And Data Engineering, VOL. 27, NO. 8, AUGUST 2015.

[5] Ch. Nanda Krishna, Dr. P. Vidya Sagar, Dr. Nageswara Rao Moparthi, "Sentiment Analysis of Top Colleges", 4th International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB-18), 2018

[6] Chaturvedi, S., Mishra, V., & Mishra, N., "Sentiment analysis using machine learning for business intelligence", International Conference on Power, Control, Signals and Instrumentation Engineering IEEE (ICPCSI), 1(1), 2162–2166, 2017.

[7] Dr. S. Gomathi, Rohini, Punitha. R, "Sentiment Analysis of Google Reviews of a College", International Conference on Computing Intelligence and Data Science (ICCIDS) 2018.

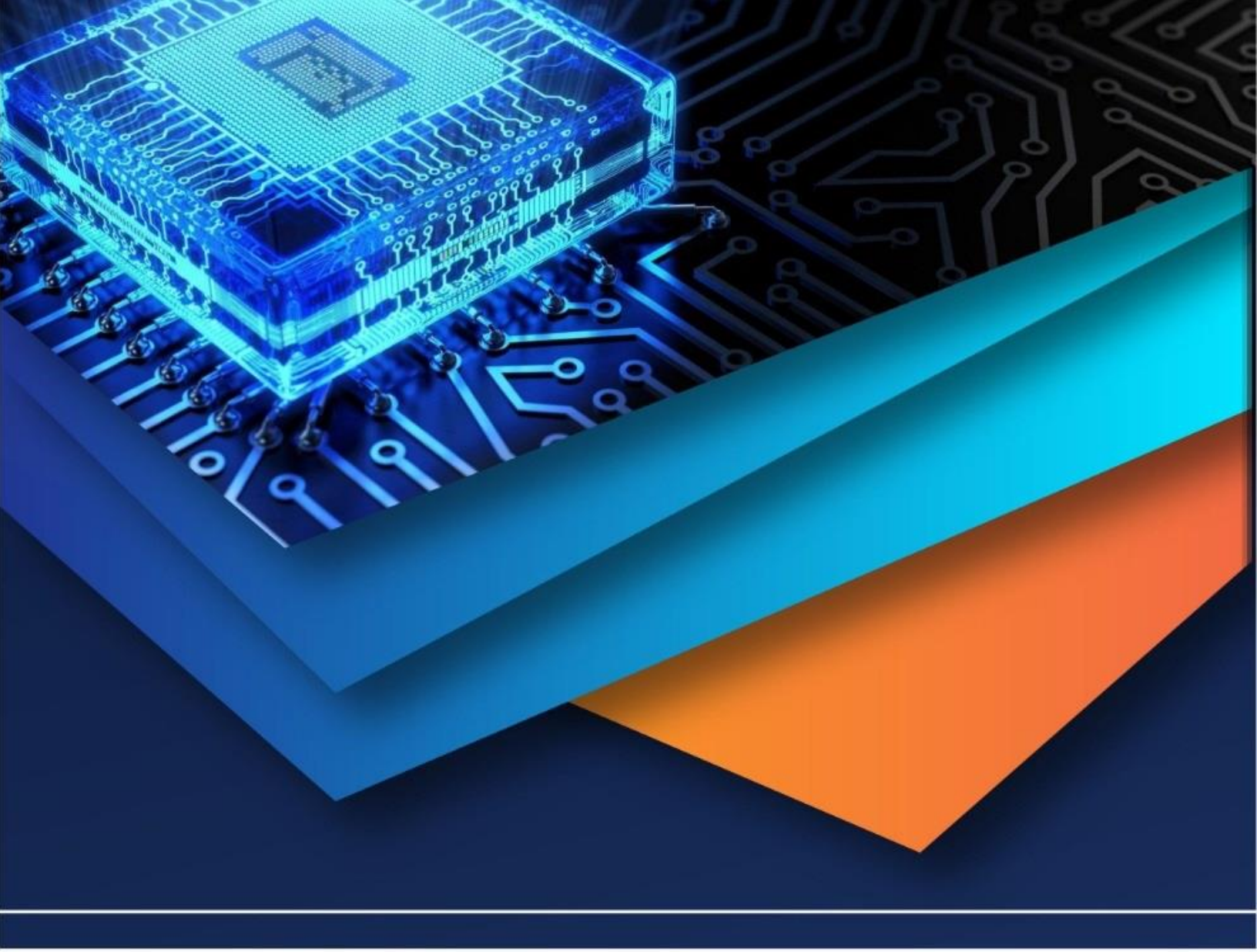
[8] Mais Yasen, Sara Tedmori, "Movies Reviews Sentiment Analysis and Classification", Jordan International joint conference on Electrical and Information Technology (JEEIT), IEEE 2019.

[9] Suci Laila Ramdhani, Rachmadita Andreswari, Muhammad Azani Hasibuan, "Sentiment Analysis of Product Reviews using Naïve Bayes Algorithm: A Case Study", The 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT), 123-127, IEEE 2018.

[10] Vivek Narayanan, Ishan Arora, Arjun Bhatia, "Fast and Accurate Sentiment Classification Using an Enhanced Naïve Bayes Model", Intelligent Data Engineering and Automated Learning – IDEAL, Springer, Vol. 8206, 2013.

[11] Nontobeko Hlongwane, Yo-Ping Huang, Li-Jen Kao, "Using Sentiment Analysis to Determine Users' Likes on Twitter", 16th Int. Conf. on Dependable, Autonomic & Secure Comp., ©2018.

[12] Dr. K. A. Waghmare, Miss. Sheetal K. Bhala, "Survey Paper on Sentiment Analysis for Tourist Reviews", 2020 International Conference on Computer Communication and Informatics (*ICCCI* -2020), Jan. 22 – 24, IEEE 2020.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY