International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

PREDICTION OF EMPLOYEE TURNOVER USING MACHINE LEARNING

Ms.M.ANITHA[1], Ms.K.PAVANI[2], Ms.K.PRIYANKA [3]

#1 Assistant professor in the Department of Master of Computer Applications in the SRK Institute of Technology, Enikepadu, Vijayawada, NTR District

#2 Assistant professor in the Department of Master of Computer Applications SRK Institute of Technology, Enikepadu, Vijayawada, NTR(DT)

#3 MCA student in the Department of Master of Computer Applications at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District.

ABSTRACT_ For a variety of reasons, employees are inclined to leave their employers. Companies who fail to retain employees experience a loss in employee productivity, are compelled to hire new workers, experience poorer morale, miss out on sales opportunities, and pay additional costs that could have been avoided if the employees had been retained in the first place. With the aid of machine learning, employee retention can be predicted. Using a decision tree method, numerical experiments are carried out in this study for real and simulated human resources datasets covering firms with small, medium, and big sized employee populations.

Our project uses Decision Tree Models to predict employees in the company and enables the Human Resource (HR) team to take the necessary action to stop them. As a result, it helps Human Resources (HR) minimise a company's loss by using Decision Tree Models to predict employees in the company and allow the HR team to take the necessary action to stop them.

1.INTRODUCTION

Decision Tree algorithms Handle both categorical and continuous data.Their interpretability, and their ability to handle non-linear relationships between variables. The disclosure or departure of an intelligent skill from a field or organisation is referred to as employee retention. Another possibility is retention, which occurs when a member of a population departs. Information about the numbers and/or percentage of employees who leave a sector and are replaced by new hires is provided by employee retention. Employers who want to investigate the causes of employee retention or calculate the cost-to-hire for financial planning may find it useful to calculate employee retention. In this study, we will discuss various retention strategies for such employees and deal with voluntary retention. We will forecast the employee's retention using various machine learning algorithms.

The disclosure or departure of an intelligent skill from a field or organisation is referred to as employee retention. Another possibility is turnover, which occurs when a member of a population departs. Because it costs money to find, hire, and train new employees, high turnover rates can be expensive for businesses. Therefore, it is crucial that businesses address the underlying causes of turnover and put strategies in place to keep their best employees. Information about the

numbers and/or percentage of employees who leave a sector and are replaced by new hires is provided by employee turnover. An organisation may suffer from low productivity, higher hiring and training costs, and a loss of institutional knowledge as a result of high employee turnover. Therefore, it is crucial that businesses address the underlying causes of employee churn and put retention plans in place.

Employers who want to investigate the causes of employee turnover or calculate the cost-to-hire for financial planning may find it useful to calculate employee turnover. Offering competitive pay and benefits packages and putting in place employee engagement programmes can also aid in keeping workers and lowering turnover rates. Organisations must put employee retention first if they want to keep a skilled and knowledgeable workforce. In this study, we will discuss various retention strategies for such employees and deal with voluntary turnover. We will forecast the employee's turnover using various machine learning algorithms. The study's objectives are to pinpoint the causes of voluntary employee departure and create a model that can precisely predict which workers are most likely to do so. According to the results, businesses can implement targeted retention strategies to lower turnover rates and keep their best employees. Recently, the development of reliable quantitative methods to extract insights from industry data has been facilitated by the prevalence of intelligent machine learning algorithms in the field of computer science. Numerous fields, including biology and the medical sciences
, transportation , political science , and many others, have demonstrated the value of supervised machine learning techniques, in which computers learn from analyses of large- scale, historical, labelled datasets. As a result of information technology advancements, researchers have also looked into numerous machine learning techniques to enhance the results of human resource (HR) management.

Several researchers have previously studied the performance evaluation of machine learning algorithms. The predictive abilities of seven different machine learning algorithms, including recently developed algorithms like Extreme Gradient Boosting [26], were notably compared by scholars, on employee turnover. Similar to this, Sikaroudi and colleagues tested various types of neural networks and induction rule techniques while simulating employee turnover using ten different data mining algorithms.

According to their research, the Extreme Gradient Boosting algorithm performed better at predicting employee turnover than other machine learning algorithms. Similar findings were found in the study by Sikaroudi and colleagues, which showed that data mining algorithms could accurately predict employee turnover using neural networks and induction rule techniques.

2.LITERATURE SURVEY

2.1 M. Stoval and N. Bontis, "Voluntary turnover: Knowledge management – Friend orfoe?", Journal of Intellectual Capital, 3(3), 303-322, 2002.

It is possible to interpret employee turnover as a loss or departure of intellectual capital from the employing organisation. In addition to increasing the costs of hiring and training new employees, this loss of intellectual capital can also result in decreased productivity and innovation. In order to prevent employee turnover, it is crucial for organisations to develop retention strategies.

2.2 J. L. Cotton and J. M. Tuttle, "Employee turnover: A meta-analysis and review with implications for research", Academy of management Review, 11(1), 55-70, 1986.

The majority of the literature on turnover divides it into two categories: voluntary and involuntary. The focus of this analysis is voluntary turnover. Job satisfaction, organisational commitment, and perceived alternative job opportunities were found to be the most reliable predictors of voluntary turnover in a meta-analytic review of voluntary turnover studies. The review also emphasised the significance of understanding and addressing voluntary turnover while taking into account both individual and organisational factors.

2.3 Najafi-Zangeneh, S.; Shams-Gharneh, N.; Arjomandi-Nezhad, A.; Zolfani, S.H. An Improved Machine Learning-Based Employees Attrition Prediction Framework with Emphasis on Feature Selection. *Mathematics* 2021, *9*, 1226.

To predict employee attrition, a three-stage system based on preprocessing, processing, and post-processing techniques was proposed [17]. The framework was tested and trained using the IBM HR employee dataset. For the dimension reduction stage, the max- out feature selection method was used. The method of logistic regression was used to forecast employee attrition. The model's output had an accuracy rating of 81%. The parameters of the framework were verified

3.PROPOSED SYSTEM

In this proposed system, Data Collection would receive the majority of our attention. Processing of Data Model Evaluation Prediction Decision Tree algorithms have the following advantages: Manage both continuous and categorical data. Their interpretability and capacity for dealing with non-linear inter-variable relationships. Due to their high accuracy and simplicity of implementation, decision tree algorithms are frequently used in a variety of industries, including finance, healthcare, and marketing. They are a popular option for data analysis because they effectively handle missing values and outliers.

3.1 IMPLEMENTATION

3.1.1 Data collection

An open-source website with many data sets, kaggle.com, is where the data was originally obtained. But after some research, it was found that Oracle had released the dataset. Even though we were aware that we couldn't find the precise dataset's release link or point of origin, we were still able to find some related links from Oracle that contained the data and provided more details about its source. The data set includes information on the satisfaction level, most recent evaluation, number of completed projects, average monthly hours, time spent with the company, work-related accidents, promotions over the previous five years, demotions, and salary. This dataset includes various factors like work accidents, promotions, and salaries that may have an impact on a worker's overall

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

job satisfaction and appears to be related to employee performance and satisfaction in a company. It might be helpful for examining employee retention rates or locating areas that could be improved to boost job satisfaction.

3.1.2 Data Cleaning

The data set must be free of any flaws that might obstruct testing or, more seriously, result in inadequate analysis. Effective solutions must be found for these flaws or issues brought on by redundant records, missing values, or loss of dimension. Bad data will therefore be removed in this step, and missing data will be added. We need to cut out any extraneous information and possibly add any that is missing from the information we currently have, which is a comprehensive general information.

```
satisfaction_level          False
last_evaluation             False
number_project              False
average_montly_hours        False
time_spend_company          False
Work_accident               False
left                        False
promotion_last_5years       False
sales                       False
salary                      False
dtype: bool
```

Fig 1. Data Cleaning

Since the data set does not contain any missing values, handling null values is a crucial step in the data cleaning process. because it makes it easier to handle issues that arise during subsequent procedures. Null values and missing values can be handled in a number of different ways. The entire data set can be eliminated, or missing values can be substituted using mean, median, or regression techniques.

3.1.3 Exploratory Dataset:

A lot of information must first be learned in order to better comprehend and explore the information present in the data. We can better understand the information in the data by making the data visually appealing. Making more informed decisions is possible with the aid of data visualisation, which can help to spot patterns, trends, and outliers that may not be immediately obvious from raw data. Furthermore, it can improve accessibility and communication with others for complex data sets.

```
Data columns (total 10 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   satisfaction_level     14999 non-null  float64
 1   last_evaluation        14999 non-null  float64
 2   number_project         14999 non-null  int64
 3   average_montly_hours   14999 non-null  int64
 4   time_spend_company     14999 non-null  int64
 5   Work_accident          14999 non-null  int64
 6   left                   14999 non-null  int64
 7   promotion_last_5years  14999 non-null  int64
 8   sales                  14999 non-null  object
 9   salary                 14999 non-null  object
dtypes: float64(2), int64(6), object(2)
memory usage: 1.1+ MB
```

Fig 2. Exploratory Dataset

The insurance data is made up with 10 features out of 10 features only 8 are numerical features and remaining all are categorical features.

3.1.4 Data transformation:

In this stage, data will be arranged or managed to make it useful for achieving the specified objective. Data preparation or preprocessing are other names for data transformation. The same thing goes by many different names. In order for your machine learning algorithm to effectively

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

use your data, it ensures that it is clean. Your AI can't make accurate predictions without data transformation. Data transformation involves a number of methods, including handling missing values, feature scaling, and normalisation. It is an essential step in the machine learning pipeline that enhances the quality of the data and lowers noise to help the model perform better.

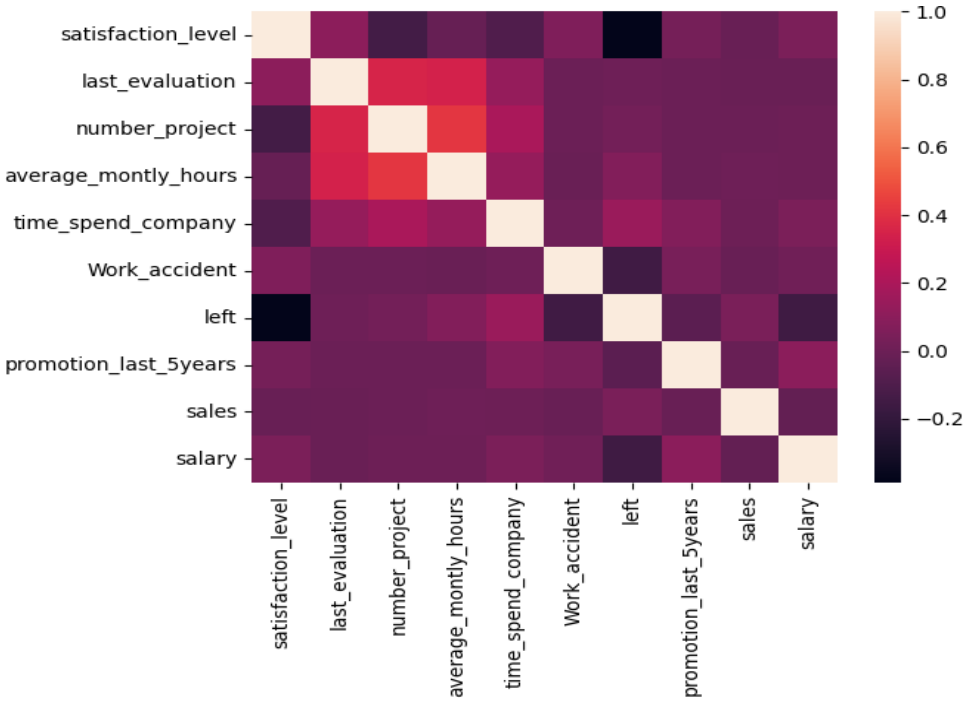

Fig 3. Relation Between Independent And Dependent Variables.

We must utilise supervised machine learning techniques to forecast the result. When we have a small amount of data, we can sometimes work with it easily, but as the data volume grows, it gets harder to identify predictors or variables. When it comes to this situation, using all the data can frequently be detrimental, which has an impact on both the computational resources and the model's accuracy. As we examine the relationship between dependent and independent features and then choose the features that are crucial for prediction, the idea of correlation enters the picture. The above figure illustrates the connection between each feature and how they correlate with one another.

3.1.5 Data Exploration:

The necessary data is extracted from the initially loaded data using Pandas data frames. Data is thoroughly examined, information is found, and then it is drawn to a conclusion to produce the report. 36 Line graphs are used to present the data so that the user can analyse it using matplotlib and seaborn. The final graph displays all of the line graphs for easier comparison. The data is displayed on different graphs to show trends in the various dataset values obtained using data visualisation before being combined into one big line graph for analysis and comparison with the same date.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

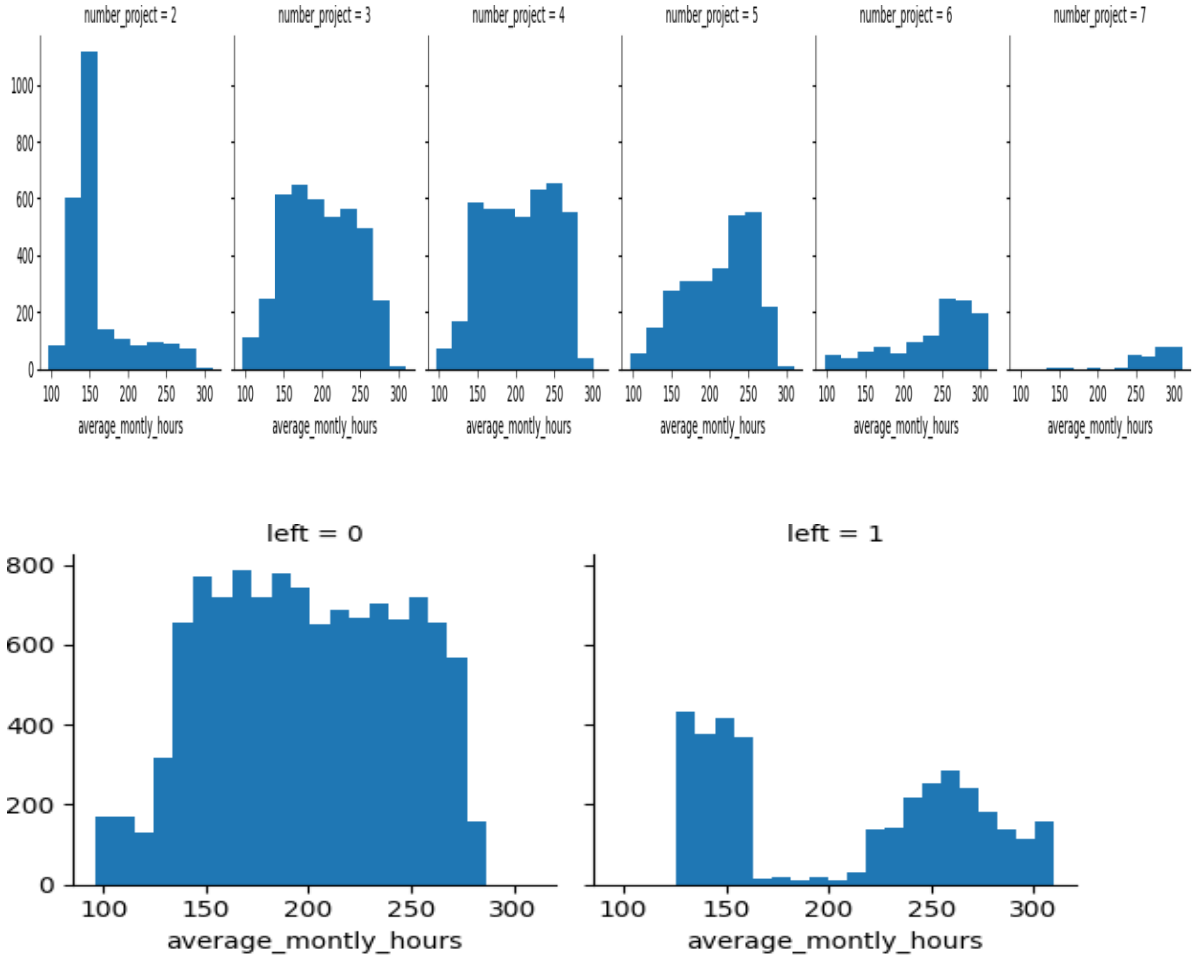

Fig 4. showing the pairwise relationships between the categorical features that are present in the dataset.

4.RESULTS AND DISCUSSION

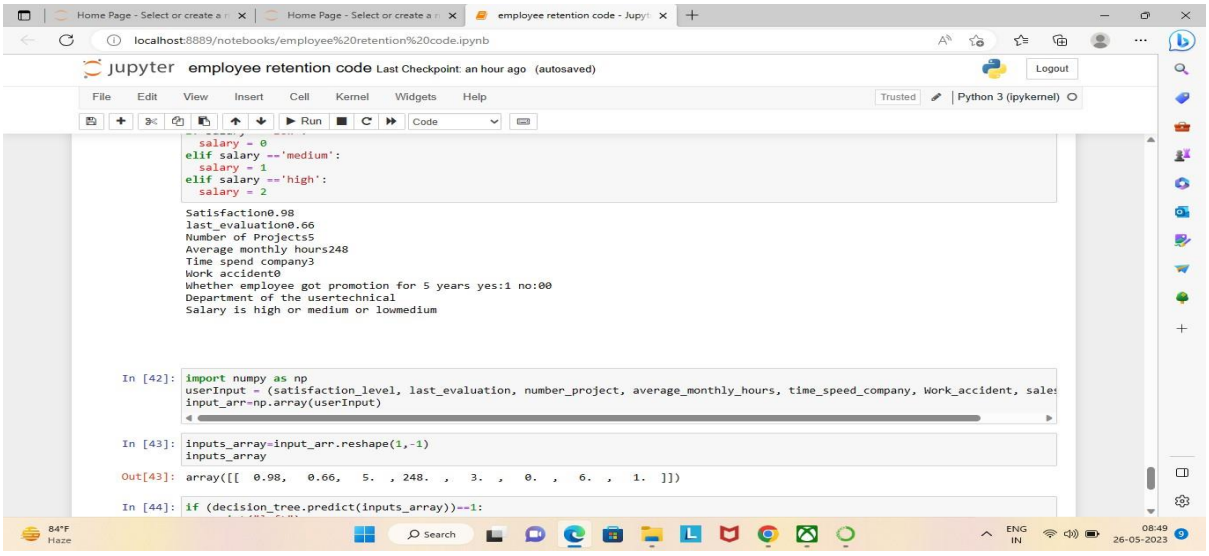

Fig 5: taking the columns data as input values

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

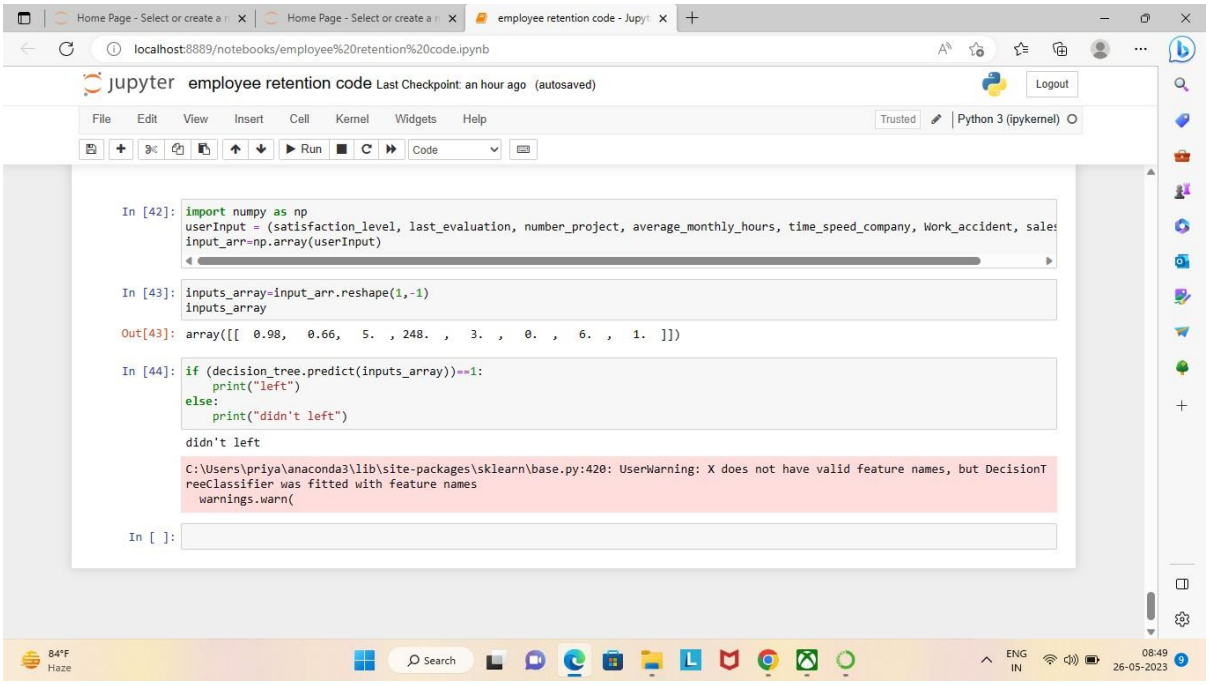

Fig 6: Represnting output as didn't left based on users input.

5.CONCLUSION

The model predicts possible rates of employee departure. HR might be able to preserve the resources using this. Offering the identified workers promotions, pay raises, or other incentives may assist HR keep them from quitting. In the long run, this can spare the business the cash and resources needed for hiring and integrating new hires. By lowering the employment process's travel and paper waste-related carbon footprint, resource saving can also help the environment. Implementing retention techniques therefore improves sustainability efforts while also enhancing the bottom line of the business.

REFERENCES:

1. M. Stoval and N. Bounties, "Voluntary turnover: Knowledge management – Friend or foe?", Journal of Intellectual Capital, 3(3), 303-322, 2002.

2. J. L. Cotton and J. M. Tuttle, "Employee turnover: A meta-analysis and review with implications for research", Academy of management Review, 11(1), 55-70, 1986.

3. Najafi-Zangeneh, S.; Shams-Gharneh, N.; Arjomandi-Nezhad, A.; Zolfani, S.H. An Improved Machine Learning-Based Employees Attrition Prediction Framework with Emphasis on Feature Selection. *Mathematics* 2021, *9*, 1226.

4. Kaya, İ.E.; Korkmaz, O. Machine Learning Approach for Predicting Employee Attrition and Factors Leading to Attrition. *Cukurova Univ. J. Fac. Eng.* 2021, *36*, 913–928.

5. Joseph, R.; Udupa, S.; Jangale, S.; Kotkar, K.; Pawar, P.

Employee attrition using machine learning and depression analysis. In Proceedings of the 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 6–8 May 2021; pp. 1000–1005.

6. Bhartiya, N.; Jannu, S.; Shukla, P.; Chapaneri, R. Employee Attrition Prediction Using Classification Models. In Proceedings of the 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 29–31 March 2019.

7. Lai, H.; Hossin, M.A.; Li, J.; Wang, R.; Hosain, M.S. Examining the Relationship between COVID-19 Related Job Stress and Employees' Turnover Intention with the Moderating Role of Perceived Organisational Support: Evidence from SMEs in China. *Int. J. Environ. Res. Public Health* 2022, *19*, 3719.

8.

AUTHOR PROFILES

Ms.M.ANITHA completed her Master of Computer Applications and Masters of Technology. Currently working as an Assistant professor in the Department of Masters of Computer Applications in the SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. Her area of interest includes Machine Learning with Python and DBMS.

Ms.K.PAVANI completed her Master of Computer Applications. Currently working as an Assistant professor in the department of MCA at SRK Institute of Technology, Enikepadu, NTR (DT). His areas of interest include Artificial Intelligence and Machine Learning.

Ms.K.PRIYANKA is an MCA student in the Department of Computer Applications at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. She has Completed Degree in B.Sc (computers) from Sir C.R.R College For Women ELURU (Affiliated to Adikavi Nannaya University)Rajamahendravaram. Her areas of interest are DBMS, JavaScript, Blockchain, Machine Learning with Python, HTML, CSS, Bootstrap and Django.