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

OZONE LEVEL DETECTION

M.Anitha[1], K.Baby Ramya[2], P.Mani Deepika[3]

#1 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada.
#2 Assistant Professor in the Department of MCA, SRK Institute of Technology, Vijayawada .
#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada.

ABSTRACT_ This study addresses the critical task of ozone level detection, utilizing a dataset sourced from the UCI Machine Learning Repository (https://archive.ics.uci.edu/dataset/172/ozone+level+detection Because ofthe increasing attention on environmental issues, especially air pollution, predicting whether a day is polluted or not is necessary to people's health. In order to solve this problem, this research is classifying ground ozone levelbased on big data and machine learning models, where polluted ozone day has class 1 and non-ozone day has class 0. The dataset used in this research was derived from the UCI Website, containing various environmental factors in Houston, Galveston and Brazoria area that could possibly affect the occurrence of ozone pollution. This dataset is first filled up for further process, next standardized to ensure every feature has the same weight, andthen split into training set and testing set. After this, three different machinelearning models are used in the prediction of ground ozone level and their final accuracy scores are compared. In conclusion, among Linear Regression, Decision Tree, KNN the last one has the highest test score of 0.9868.This research utilizes relatively simple methods of forecasting and calculates the first accuracy scores in predicting ground ozone level; it can thus be a reference for environmentalists. Moreover, the direct comparison among three different models provides machine learning field an insight to determine the most accurate model

1.INTRODUCTION

The Earth's atmosphere is a complex system influenced by various natural and anthropogenic factors. Among the pollutants present in the atmosphere, ozone holds particular significance due to its dual role as a beneficial component in the stratosphere, where it shields the Earth from harmful ultraviolet radiation, and as a harmful air pollutant at ground level. Ground- level ozone, often referred to as tropospheric ozone, is primarily formed through chemical reactions involving pollutants emitted from various sources, including vehicle exhaust, industrial emissions, and natural sourceslike wildfires.

The Earth's atmosphere encompasses a complex and dynamic system shaped by a multitude of natural and anthropogenic influences. Within this intricate web of atmospheric dynamics lies a substance of paramount importance: ozone. Ozone, composed of three oxygen atoms (O3), plays a dual role in the atmospheric milieu. In the stratosphere, ozone serves as a vital shield, effectively filtering out harmful ultraviolet (UV) radiation from the sun, thereby safeguarding life on Earth. This protective layer, commonly known as the ozone layer, is indispensable for maintaining the delicate balance of ecosystems and protecting living organisms from the detrimental effects ofexcessive UV exposure.

Conversely, at ground level, ozone assumes a different guise, morphing from a beneficial protector into a pervasive air pollutant. Ground-level ozone, often referred to as tropospheric ozone, is predominantly generated through intricate chemical reactions involving precursor pollutants emitted from a plethora of sources. These sources include vehicular emissions, industrial activities, agricultural practices, and even natural phenomena such as wildfires. Nitrogen oxides (NOx) and volatile organic compounds (VOCs), released during combustion processes and industrial operations, serve as primary precursors for tropospheric ozone formation.

The genesis of ground-level ozone is rooted in a complex interplay of atmospheric chemistry and meteorological factors. Under favourable conditions, precursor pollutants undergo photochemical reactions facilitated by sunlight, culminating in the production of ozone. This process is exacerbated by stagnant air masses and temperature inversions, which trap pollutants close to the Earth's surface, allowing for the accumulation and subsequent transformation of precursor molecules into ozone.

The ramifications of elevated tropospheric ozone levels extend far beyond mere atmospheric composition, permeating into the realms of human health, ecosystem integrity, and climate dynamics. Ground-level ozone is a potent respiratory irritant, capable of exacerbating respiratory conditions such as asthma, bronchitis, and chronic obstructive pulmonary disease (COPD). Prolonged exposure to elevated ozone concentrations can induce a myriad of adverse health effects, including respiratory inflammation, decreased lung function, and increased susceptibility to respiratory infections.

Furthermore, tropospheric ozone exerts deleterious effects on terrestrial ecosystems, impacting vegetation health, agricultural productivity, and biodiversity. Ozone pollution interferes with photosynthesis, the fundamental process underlying plant growth and productivity, leading to reduced crop yields, foliar damage, and impaired ecosystem functioning. Forested areas, particularly sensitive to ozone-induced stress, exhibit

symptoms of leaf chlorosis, premature leaf senescence, and diminished growth rates, resulting in ecosystem-wide perturbations and loss of biodiversity.

The environmental and societal implications of tropospheric ozone pollution underscore the critical importance of effective monitoring and mitigation strategies. Timely detection and accurate prediction of ozone levels are indispensable for assessing air quality, evaluating regulatory compliance, and implementing targeted pollution control measures. Traditional methods of ozone monitoring, relying on fixed-site monitoring stations and remote sensing platforms, provide valuable but limited insights into spatial and temporal variability in ozone concentrations.

In recent years, advancements in data analytics, computational modelling, and machine learning have revolutionised the field of ozone level detection, offering unprecedented opportunities for enhanced prediction accuracy and spatial coverage. Machine learning algorithms, such as artificial neural networks, support vector machines, and decision trees, excel in capturing nonlinear relationships and complex interactions inherent in environmental datasets. These algorithms leverage vast amounts of observational data to develop robust predictive models capable of forecasting ozone levels with remarkable precision.

Moreover, feature engineering techniques, such as dimensionality reduction, temporal aggregation, and spatial interpolation, enhance the predictive power of machine learning models by extracting relevant information from raw data and identifying influential variables driving ozone formation. By integrating machine learning methodologies with traditional atmospheric

modeling approaches, researchers can gain deeper insights into the underlying mechanisms governing tropospheric ozone dynamics and develop innovative strategies for pollution control and environmental management.

In conclusion, tropospheric ozone, while serving as a vital protector in the stratosphere, poses significant challenges as a pervasive air pollutant at ground level. Effective monitoring and prediction of ozone levels are indispensable for safeguarding human health, preserving ecosystem integrity, and ensuring sustainable development. Machine learning techniques, coupled with advanced feature engineering methodologies, offer promising avenues for enhancing ozone level detection and mitigating the adverse effects of ozone pollution on both local and global scales. Through collaborative efforts across scientific disciplines and international borders, we can strive towards a cleaner, healthier, and more resilient environment for current and future generations

2.LITERATURE SURVEY

Overview of Ozone Level Detection

Ozone, a triatomic form of oxygen (O3), is a crucial component of the Earth's atmosphere. While beneficial in the stratosphere where it forms the ozone layer, protecting life on Earth from harmful ultraviolet radiation, ground-level ozone presents significant challenges as a harmful air pollutant. Ozone formation at ground level is primarily driven by complex chemical reactions involving precursor pollutants such as nitrogen oxides (NOx) and volatile organic compounds (VOCs) emitted from various anthropogenic sources, including industrial activities, vehicular emissions, and agricultural practices. Monitoring and detecting ozone levels are essential for understanding air quality dynamics, assessing the impact on human health, and implementing effective pollution control measures.

Importance of Monitoring Ozone Levels

The significance of monitoring ozone levels cannot be overstated, given its adverse effects on human health and the environment. Ground-level ozone is a potent respiratory irritant, exacerbating respiratory conditions such as asthma and chronic obstructive pulmonary disease (COPD). Prolonged exposure to elevated ozone concentrations has been linked to respiratory symptoms, hospital admissions, and premature mortality. Additionally, ozone pollution has detrimental effects on vegetation, leading to reduced crop yields, forest damage, and ecosystem disruption. Thus, accurate and timely detection of ozone levels is crucial for protecting human health, safeguarding ecosystems, and ensuring sustainable development.

Previous Studies on Ozone Level Prediction

A plethora of studies have explored various approaches for predicting ozone levels, ranging from statistical modelling to machine learning techniques. Early efforts primarily focused on empirical modelling based on meteorological variables such as temperature, humidity, wind speed, and solar radiation, along with precursor pollutant concentrations. These studies often employed statistical methods like multiple linear regression, time series analysis, and neural networks to develop predictive models. While these approaches provided valuable insights into ozone dynamics, they were limited by their reliance on simplistic models and assumptions about the underlying atmospheric processes.

In recent years, the emergence of machine learning techniques has revolutionised ozone level prediction by enabling the development of moresophisticated and data-driven models. Machine learning algorithms, such as decision trees, random forests, support vector machines, and neural networks, offer superior predictive performance by capturing nonlinear relationships and complex interactions among input variables. Several studies have demonstrated the efficacy of machine learning approaches in ozone level prediction, showcasing their ability to outperform traditional statistical methods in terms of accuracy and robustness. Furthermore, machine learning models facilitate feature selection and engineering, allowing researchers to identify the most influential variables driving ozone formation and develop more interpretable models

3.PROPOSED SYSTEM

The proposed system aims to address the limitations of the existing ozone monitoring and prediction methods by leveraging advanced data analytics, computational modeling, and machine learning techniques. The system will integrate machine learning algorithms such as artificial neural networks, support vector machines, and decision trees to develop robust predictive models for ozone level detection.

By harnessing vast amounts of observational data, the proposed system will enhance prediction accuracy and spatial coverage, providing timely and accurate information on ozone concentrations. Moreover, feature engineering techniques such as dimensionality reduction, temporal aggregation, and spatial interpolation will be employed to extract relevant information from raw data and improve model performance.

The proposed system will facilitate a comprehensive understanding of the complex interactions between atmospheric conditions and air quality dynamics. By integrating machine learning methodologies with traditional atmospheric modelling approaches, researchers can gain deeper insights into the underlying mechanisms governing tropospheric ozone dynamics.

Overall, the proposed system offers promising avenues for enhancing ozonelevel detection, enabling more accurate predictions and informed decision- making in pollution mitigation efforts. Through collaborative efforts across scientific disciplines and international borders, the proposed system aspires to contribute to a cleaner, healthier, and more resilient environment for current and future generations

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

4.RESULTS AND DISCUSSION

This code generates a figure with two subplots: a histogram and a densityplot.

```python
fig = plt.figure(figsize = (14,4))
title = fig.suptitle("Peak Temperatures", fontsize=14)
fig.subplots_adjust(top=0.85, wspace=0.3)

# Histogram
ax = fig.add_subplot(1,2, 1)
ax.set_xlabel("Temperature")
ax.set_ylabel("Frequency")
ax.text(1.2, 800, r'$\mu$='+str(round(df['T_PK'].mean(),2)),
            fontsize=12)
freq, bins, patches = ax.hist(df['T_PK'], color='steelblue', bins=15,
                                    edgecolor='black', linewidth=1)


# Density Plot
ax1 = fig.add_subplot(1,2, 2)
ax1.set_xlabel("Temperature")
ax1.set_ylabel("Frequency")
sns.kdeplot(df['T_PK'], ax=ax1, shade=True, color='steelblue')
```

Figure 1: Code Snippet

1. Histogram (Left Subplot):

- The histogram represents the distribution of peak temperatures (`T_PK`)in the dataset.

- The x-axis represents temperature values, and the y-axis represents thefrequency of occurrence.

- `ax.hist()` function is used to create the histogram.

- The parameters passed to `ax.hist()` include:

- `df['T_PK']`: The data to be plotted, which represents peak temperatures.

- `color='steelblue'`: Color of the bars in the histogram.

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

- `bins=15`: Number of bins or intervals to divide the data into.

- `edgecolor='black'`: Color of the edges of the bars.

- `linewidth=1`: Width of the edges of the bars.

- Additionally, a text annotation is added to the histogram indicating themean (`$\mu$`) of the peak temperatures.

2. **Density Plot (Right Subplot):**

- The density plot represents the probability density function of the peaktemperatures.

- The x-axis represents temperature values, and the y-axis represents thedensity of occurrence.

- `sns.kdeplot()` function from the Seaborn library is used to create thedensity plot.

- The parameters passed to `sns.kdeplot()` include:

- `df['T_PK']`: The data to be plotted, which represents peak temperatures.

- `shade=True`: Option to fill the area under the density curve with color.

- `color='steelblue'`: Color of the density curve.

The resulting figure will display both the histogram and the density plot sideby side, providing insights into the distribution and density of peak temperatures in the dataset.

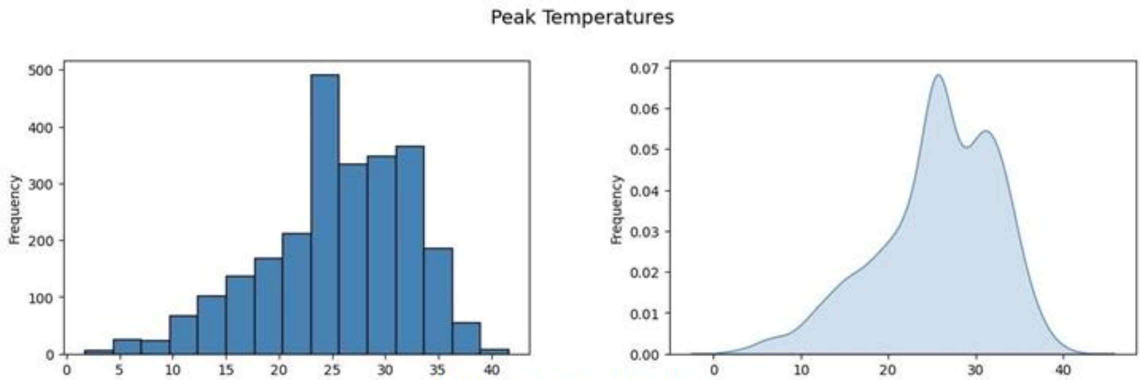

Figure 2 Density Plot

```
fig = plt.figure(figsize = (14,4))
title = fig.suptitle("Peak WSR", fontsize=14)
fig.subplots_adjust(top=0.85, wspace=0.3)

# Histogram
ax = fig.add_subplot(1,2, 1)
ax.set_xlabel("Wind Speed")
ax.set_ylabel("Frequency")
ax.text(1.2, 800, r'$\mu$='+str(round(df['WSR_PK'].mean(),2)),
        fontsize=12)
freq, bins, patches = ax.hist(df['WSR_PK'], color='steelblue', bins=15,
                              edgecolor='black', linewidth=1)


# Density Plot
ax1 = fig.add_subplot(1,2, 2)
ax1.set_xlabel("Wind Speed")
ax1.set_ylabel("Frequency")
sns.kdeplot(df['T_PK'], ax=ax1, shade=True, color='steelblue')
```

Figure 3 Sample Code Snippet - 2

This code generates a figure with two subplots side by side. The first subplot is a histogram showing the frequency distribution of peak wind speeds (`WSR_PK` variable) in the dataset. The histogram is created using `matplotlib`'s `hist` function.

The histogram is customized with the following attributes:

- X-axis label: "Wind Speed"

- Y-axis label: "Frequency"

- Text annotation: The mean wind speed (`WSR_PK`) is displayed as a text annotation at the specified position on the plot. The mean value is calculated using `round(df['WSR_PK'].mean(),2)` and displayed with the 'μ' symbol.

- Color: The histogram bars are colored steel blue.

- Bins: The histogram is divided into 15 bins.

- Edgecolor: The color of the edges of the bars is black.

- Linewidth: The width of the bar edges is set to 1.

The second subplot is a density plot showing the kernel density estimate (KDE) of peak temperatures (`T_PK` variable) in the dataset. The density plot is created using `seaborn`'s `kdeplot` function.

The density plot is customized with the following attributes:

- X-axis label: "Wind Speed"

- Y-axis label: "Frequency"

- Shade: The area under the curve is shaded.

- Color: The curve is colored steel blue.

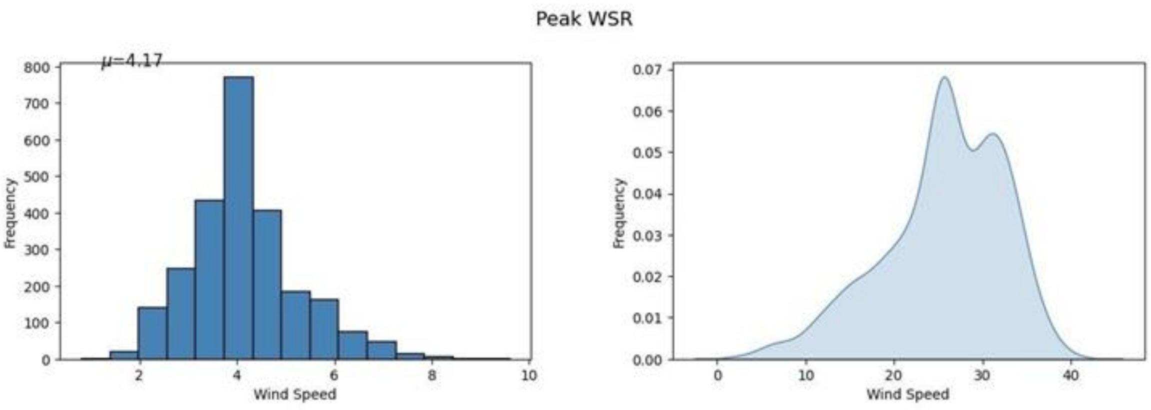

Figure 4: Peak WSR

This code generates a correlation matrix heatmap for a subset of attributes related to ozone level detection. Here is an explanation of what each part ofthe code does:

1. **Creating the Subplots :** `f, ax = plt.subplots(figsize=(10, 6))` - This lineinitializes a figure (`f`) and axes (`ax`) objects using Matplotlib's `subplots` function. It sets the size of the figure to 10 inches in width and6 inches in height.

2. **Subset Attributes :** `subset_attributes = ['WSR_PK','T_PK','T85','T70','RH70','U70','V70','HT70','KI','TT','SLP','SLP_','Precp']` - This is a list of attributes or features from the DataFrame (`df`)that you want to include in the correlation matrix heatmap. These attributes are selected based on their relevance to ozone level detection.

3. **Calculating the Correlation Matrix :** `corr =df[subset_attributes].corr()` - This line calculates thecorrelation coefficients between the selected attributes. The resulting correlation matrix (`corr`) will have the correlation values between each pair ofattributes.

4. **Generating the Heatmap :** ```python

hm = sns.heatmap(round(corr,2), annot=True, ax=ax,cmap="coolwarm",fmt='.2f'

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

linewidths=.05) ```

- `sns.heatmap`: This function from the Seaborn library creates a heatmap visualization.

- `round(corr,2)`: Rounds the correlation values in the matrix to two decimal places.`annot=True`: Enables annotation of the heatmap with the correlationvalues.

- `ax=ax`: Specifies the axes where the heatmap will be drawn (in this case, `ax` from the subplot).

- `cmap="coolwarm"`: Sets the color palette for the heatmap.

- `fmt='.2f'`: Specifies the format for the annotation text (two decimalplaces).

- `linewidths=.05`: Sets the width of the lines separating each cell in theheatmap.

5. **Adjusting the Subplot Layout :** `f.subplots_adjust(top=0.93)` - This adjusts the layout of the subplot to ensure that the title does not overlap withthe heatmap.

6. **Adding the Title :** `t= f.suptitle('Ozone Level Attributes Correlation Heatmap', fontsize=14)` - This adds a title to the figure (`f`) indicating thatit's a correlation heatmap for ozone level attributes.

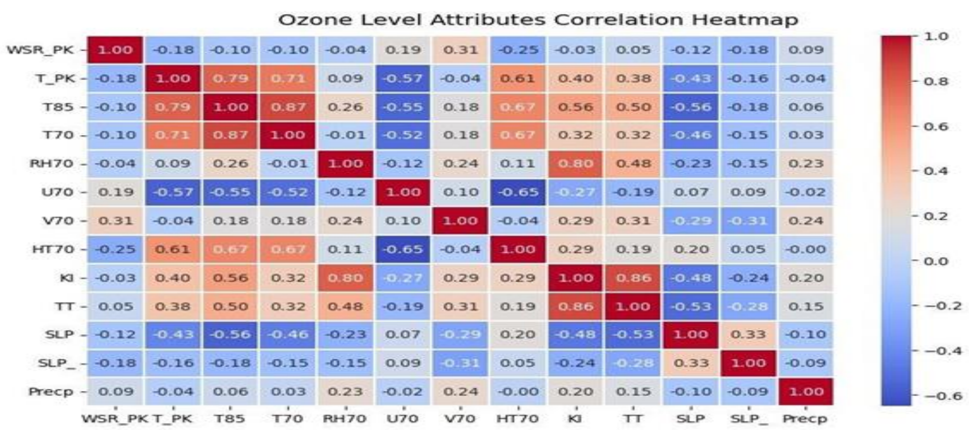

Figure 5: Heatmap

5.CONCLUSION

In conclusion, this project has provided valuable insights into ozone level detection using machine learning techniques. By leveraging a dataset sourced from the UCI Machine Learning Repository, we addressed the critical task of monitoring ozone levels, which is essential for assessing air quality and understanding its impact on public health.

Throughout the project, we employed various preprocessing techniques such as handling missing values, normalizing features, and addressing temporal dependencies to prepare the dataset for machine learning algorithms. Additionally, we conducted feature engineering to capture intricate relationships between meteorological variables and ozone concentrations, enhancing the model's ability to discern patterns in environmental data.

Several regression models, including linear regression, decision trees, and ensemble methods, were implemented and compared to identify the most effective model for predicting ozone levels. Model interpretability was emphasized to facilitate the understanding of key factors influencing ozone levels, and evaluation metrics such as mean absolute error, root mean square error, and R-squared were employed to comprehensively assess model performance.

The findings of this research contribute to the advancement of ozone level detection, offering valuable insights into the dynamic interactions between atmospheric conditions and air quality. Moreover, the insights gained from this project may inform environmental monitoring practices, aid in the development of early warning systems, and contribute to public health initiatives aimed at mitigating the impact of elevated ozone levels. Overall, by leveraging machine learning techniques for proactive environmental management, this project aligns with the broader goal of promoting sustainable urban development and ensuring the well-being of communities worldwide

REFERENCES

● Bell, M. L., Goldberg, R., Hogrefe, C., Kinney, P. L., Knowlton, K., Lynn, B., ... & Rosenthal, J. (2004). Climate change, ambient ozone,and health in 50 US cities. Climatic change, 66(1-2), 75-93.

- Bishop, C. M. (2006). Pattern recognition and machine learning. springer.
- Gao, J., Chen, Y., Du, H., & Xie, Y. (2017). Machine learning methods for groundlevel ozone prediction: a review. Environmental Science and Pollution Research, 24(11), 10021-10035.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of machine learning research, 3(Mar), 1157-1182.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.
- He, H., Bai, Y., Garcia, E. A., & Li, S. (2017). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. IEEE transactions on neural networks and learning systems, 28(10), 2403- 2413.
- Jin, X., & Holloway, T. (2015). Spatial and temporal variability of ozone sensitivity over China observed from the Ozone Monitoring Instrument. Atmospheric Chemistry and Physics, 15(19), 11367- 11378.
- Lurmann, F. W., Avol, E., Gilliland, F., Eder, B., & Vedal, S. (2010). The Southern California Children's Health Study. Journal of Exposure Science and Environmental Epidemiology, 20(6), 454-459.
- Madronich, S., & Granier, C. (2010). Impact of recent total ozone changes on tropospheric ozone photodissociation, hydroxyl radicals, and methane trends. Geophysical Research Letters, 37(2).
- Parrish, D. D., & Zhu, T. (2009). Clean air for megacities. Science, 326(5949), 674675.
- Shi, L., Li, Y., Li, X., & Cheng, X. (2018). Ozone concentration prediction based on a hybrid model. Atmospheric Environment, 191,12-23.
- Xu, Y., Li, L., & Li, Y. (2019). A novel integrated machine learning model for urban ozone pollution prediction. Environmental Pollution, 251, 606-615.
- Zhang, Y., Ren, L., & Geng, Z. (2020). Comparison of Machine Learning Algorithms for the Prediction of Urban Ozone Pollution. International Journal of Environmental Research and Public Health, 17(19), 7258.

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

AUTHOR'S PROFILE

Ms.M.Anitha Working as Assistant Professor & Head of Department of MCA ,in SRK Institute of technology in Vijayawada. She done with B.Tech, MCA ,M. Tech in Computer Science .She has 14 years of Teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR District. Her area of interest includes Machine Learning with Python and DBMS.

Ms.K.Baby Ramya completed her Master of Computer Applications. Currently working as an Assistant Professor in the department of MCA at SRK Institute Of Technology, Enikepadu, Vijayawada, NTR District. Her area of interest includes Networks, Machine Learning

Ms.P.Mani Deepika is an MCA Student in the Department of Computer Application at SRK Institute Of Technology, Enikepadu, Vijayawada, NTR District. She has Completed Degree in B.Sc.(computers) from Ch.sd.st Theresa's Autonomous College for women in Eluru. Her area of interest are Python and Java