



## **AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY**

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

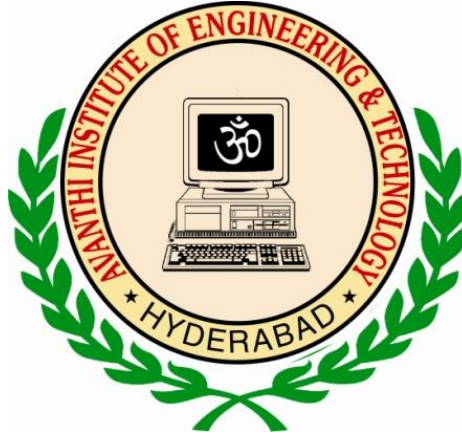
NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanathi@gmail.com](mailto:principal.avanathi@gmail.com)

---

## **DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**



## **WEB & SOCIAL MEDIA ANALYTICS LAB MANUAL**

Regulation : R18/JNTUH

Academic Year : 2023-24

Prepared By

**Shaik Subhan Ali**  
**Assistant Professor**

## **COMPUTER SCIENCE AND ENGINEERING**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

Avanathi Institute of Engineering and Technology



# **AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY**

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

## **VISION AND MISSION OF THE INSTITUTION**

### **VISION**

To develop highly skilled professionals with ethics & human values

### **MISSION**

1. To provide high-quality education along with professional training and exposure to the workplace.
2. To encourage a professional mindset that goes beyond academic achievement.
3. To promote holistic education among Department students by means of integrated pedagogy and scholarly mentoring for excellence in both personal and professional domains.
4. To consistently enhance the teaching and learning procedures in order to prepare students for successful careers in business or overseas or in further education.
5. To carefully prepare students to be Globally employable professionals who will meet societal demands and contribute to the nation's technological advancement through their research and innovative talents.

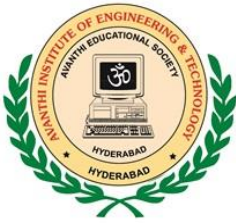
## **VISION AND MISSION OF CSE DEPARTMENT**

### **VISION**

To become a center of excellence the computer science and information technology discipline with a strong research and teaching environment.

### **MISSION**

1. To provide qualitative education and generate new knowledge by engaging in cutting edge research and by offering state of the art undergraduate, post graduate, leading careers as computer professional in the widely diversified of industry, government and academia.
2. To promote a teaching and learning process that yields advancements in state of art in computer science and engineering in integration of research result and innovative into other scientific discipline leading to new products.
3. To harness human capital for sustainable competitive edge and social relevance by including the philosophy of continuous learning and innovation in computer science and engineering.



# AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

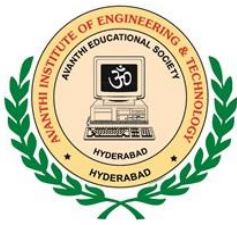
### PROGRAM EDUCATIONAL OBJECTIVES (PEOS):

A graduate of the Computer Science and Engineering Program should:

<b>PEO1</b>	<b>Program Educational Objective1: (PEO1)</b> The Graduates will provide solutions to difficult and challenging issues in their profession by applying computer science and engineering theory and principles.
<b>PEO2</b>	<b>Program Educational Objective2 :( PEO2)</b> The Graduates have successful careers in computer science and engineering fields or will be able to successfully pursue advanced degrees.
<b>PEO3</b>	<b>Program Educational Objective2 :( PEO3)</b> The Graduates will communicate effectively, work collaboratively and exhibit high levels of Professionalism, moral and ethical responsibility.
<b>PEO4</b>	<b>Program Educational Objective2 :( PEO4)</b> The Graduates will develop the ability to understand and analyse Engineering issues in a broader perspective with ethical responsibility towards sustainable development.

### PROGRAM OUTCOMES (POS):

<b>PO1</b>	<b>Engineering knowledge:</b> Apply the knowledge of mathematics, science, engineering Fundamentals and an engineering specialization to the solution of complex engineering problems.
<b>PO2</b>	<b>Problem analysis:</b> Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences
<b>PO3</b>	<b>Design/development of solutions:</b> Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations
<b>PO4</b>	<b>Conduct investigations of complex problems:</b> Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
<b>PO5</b>	<b>Modern tool usage:</b> Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations
<b>PO6</b>	<b>The engineer and society:</b> Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice
<b>PO7</b>	<b>Environment and sustainability:</b> Understand the impact of the professional engineering Solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC "B++" Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

<b>PO8</b>	<b>Ethics:</b> Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
<b>PO9</b>	<b>Individual and team work:</b> Function effectively as an individual, and as a member or leader In diverse teams, and in multi-disciplinary settings.
<b>PO10</b>	<b>Communication:</b> Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
<b>PO11</b>	<b>Project management and finance:</b> Demonstrate knowledge and understanding of the Engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments
<b>PO12</b>	<b>Life-long learning:</b> Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

### PROGRAM SPECIFIC OUTCOMES(PSOS):

<b>PSO1</b>	<b>Problem Solving Skills</b> – Graduate will be able to apply computational techniques and software principles to solve complex engineering problems pertaining to software engineering.
<b>PSO2</b>	<b>Professional Skills</b> – Graduate will be able to think critically, communicate effectively, and collaborate in teams through participation in co and extra-curricular activities
<b>PSO3</b>	<b>Successful Career</b> – Graduates will possess a solid foundation in computer science and engineering that will enable them to grow in their profession and pursue lifelong learning through post-graduation and professional development.



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanathi@gmail.com](mailto:principal.avanathi@gmail.com)

### Course Objectives

- To provide hands-on experience on web technologies.
- To develop client-server application using web technologies
- To introduce server-side programming with Java servlets and JSP
- To understand the various phases in the design of a compiler
- To understand the design of top-down and bottom-up parsers.
- To understand syntax directed translation schemes.
- To introduce lex and yacc tools.

### CO-PO & PSO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	2	-	-	1	-	-	-	-	-	-	-	-	-	-	-
CO2	2	-	-	1	-	-	-	-	-	-	-	-	1	-	-
CO3	2	1	-	1	-	-	-	-	-	-	-	-	1	1	-
CO4	2	2	2	1	-	-	-	-	-	-	-	-	2	2	1



# AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanathi@gmail.com](mailto:principal.avanathi@gmail.com)

## B. TECH CSE- 4-1

### COMPILER DESIGN LAB

S.No	List Of Experiments
1	Preprocessing text document using NLTK of Python a) Stopword elimination b) Stemming c) Lemmatization d) POS tagging e) Lexical analysis
2	Sentiment analysis on customer review on products
3	Web analytics a) Web usage data (web server log data, clickstream analysis) b) Hyperlink data
4	Search engine optimization- implement spamdexing
5	Use Google analytics tools to implement the following a) Conversion Statistics b) Visitor Profiles
6	Use Google analytics tools to implement the Traffic Sources.



# AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

## WEB AND SOCIAL MEDIA ANALYTICS LAB

B.Tech. IV Year I Sem.

L T P C

0 0 2 1

**Course Objectives:** Exposure to various web and social media analytic techniques.

### Course Outcomes:

1. Knowledge on decision support systems.
2. Apply natural language processing concepts on text analytics.
3. Understand sentiment analysis.
4. Knowledge on search engine optimization and web analytics.

### List of Experiments

1. Preprocessing text document using NLTK of Python
  - a) Stopword elimination
  - b) Stemming
  - c) Lemmatization
  - d) POS tagging
  - e) Lexical analysis
2. Sentiment analysis on customer review on products
3. Web analytics
  - a) Web usage data (web server log data, clickstream analysis)
  - b) Hyperlink data
4. Search engine optimization- implement spamdexing
5. Use Google analytics tools to implement the following
  - a) Conversion Statistics
  - b) Visitor Profiles
6. Use Google analytics tools to implement the Traffic Sources.

### Resources:

1. Stanford core NLP package
2. [GOOGLE.COM/ANALYTICS](http://GOOGLE.COM/ANALYTICS)

### TEXT BOOKS:

1. Ramesh Sharda, Dursun Delen, Efraim Turban, BUSINESS INTELLIGENCE AND ANALYTICS: SYSTEMS FOR DECISION SUPPORT, Pearson Education.

### REFERENCE BOOKS:

1. RajivSabherwal, Irma Becerra- Fernandez,” Business Intelligence –Practice, Technologies and Management”, John Wiley 2011.
2. Lariss T. Moss, Shaku Atre, “Business Intelligence Roadmap”, Addison-Wesley It Service.
3. Yuli Vasiliev, “Oracle Business Intelligence: The Condensed Guide to Analysis and Reporting”, SPD Shroff, 2012.



# AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

## PROGRAM - 1

### 1. Preprocessing text document using NLTK of Python

#### a) Stopword Elimination

```
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

example_sent = """This is a program for stop list,
so filter the words students."""

stop_words = set(stopwords.words('english'))

word_tokens = word_tokenize(example_sent)
# converts the words in word_tokens to lower case and then checks whether
#they are present in stop_words or not
filtered_sentence = [w for w in word_tokens if not w.lower() in stop_words]
#with no lower case conversion
filtered_sentence = []

for w in word_tokens:
    if w not in stop_words:
        filtered_sentence.append(w)

print(word_tokens)
print(filtered_sentence)
```

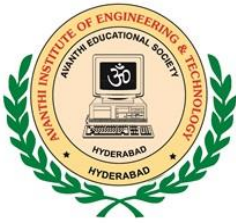
#### OUTPUT :

```
word_tokens = ['This', 'is', 'a', 'program', 'for', 'stop', 'list', ',', 'so', 'filter', 'the', 'words', 'students', '.']
filtered_sentence = ['This', 'program', 'stop', 'list', ',', 'filter', 'words', 'students', '.']
```

### VIVA Questions:

1. What are stopwords and why are they removed from text data?
2. Can you provide examples of some common stopwords?
3. How can the removal of stopwords affect the performance of a text classification model?





## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

### b) Stemming:

```
from nltk.stem import PorterStemmer

# List of words to be stemmed
e_words = ["wait", "waiting", "waited", "waits"]

# Initialize the PorterStemmer
ps = PorterStemmer()

# Stem each word in the list and print the root word
for w in e_words:
    rootWord = ps.stem(w)
    print(rootWord)
```

#### Output:

```
wait
wait
wait
wait
```

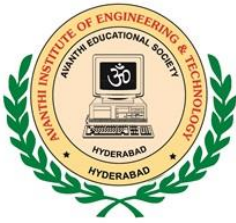
### c) Stemming Tokens from a Sentence

```
import nltk
from nltk.stem.porter import PorterStemmer

# Initialize the PorterStemmer
porter_stemmer = PorterStemmer()

# Sentence to be tokenized and stemmed
word_data = "It originated from the idea that there are readers who prefer learning new skills from the comforts of their drawing rooms."

# First, tokenize the sentence into words
nltk_tokens = nltk.word_tokenize(word_data)
```



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

# Print the actual word and its stemmed version

for w in nltk\_tokens:

```
print("Actual: %s Stem: %s" % (w, porter_stemmer.stem(w)))
```

### Output:

Actual: It Stem: it

Actual: originated Stem: origin

Actual: from Stem: from

Actual: the Stem: the

Actual: idea Stem: idea

Actual: that Stem: that

Actual: there Stem: there

Actual: are Stem: are

Actual: readers Stem: reader

Actual: who Stem: who

Actual: prefer Stem: prefer

Actual: learning Stem: learn

Actual: new Stem: new

Actual: skills Stem: skill

Actual: from Stem: from

Actual: the Stem: the

Actual: comforts Stem: comfort

Actual: of Stem: of

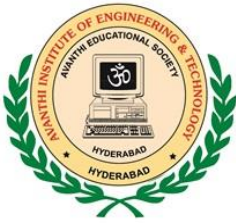
Actual: their Stem: their

Actual: drawing Stem: draw

Actual: rooms Stem: room

### VIVA Questions :

1. What is stemming and how does it differ from lemmatization?
2. Name some commonly used stemming algorithms.
3. What are the potential drawbacks of using stemming?



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

### d) Lemmatization:

```
import nltk
from nltk.stem import WordNetLemmatizer

# Download necessary NLTK data
nltk.download('punkt')
nltk.download('wordnet')

# Initialize the WordNetLemmatizer
wordnet_lemmatizer = WordNetLemmatizer()

# Sentence to be tokenized and lemmatized
word_data = "It originated from the idea that there are readers who prefer learning new skills from
the comforts of their drawing rooms."

# Tokenize the sentence into words
nltk_tokens = nltk.word_tokenize(word_data)

# Print the actual word and its lemmatized version
for w in nltk_tokens:
    print("Actual: %s Lemma: %s" % (w, wordnet_lemmatizer.lemmatize(w)))
```

### Output:

Actual: It Lemma: It

Actual: originated Lemma: originated

Actual: from Lemma: from

Actual: the Lemma: the

Actual: idea Lemma: idea

Actual: that Lemma: that

Actual: there Lemma: there

Actual: are Lemma: are



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

Actual: readers Lemma: reader

Actual: who Lemma: who

Actual: prefer Lemma: prefer

Actual: learning Lemma: learning

Actual: new Lemma: new

Actual: skills Lemma: skill

Actual: from Lemma: from

Actual: the Lemma: the

Actual: comforts Lemma: comfort

Actual: of Lemma: of

Actual: their Lemma: their

Actual: drawing Lemma: drawing

Actual: rooms Lemma: room

### VIVA Questions :

1. What is lemmatization and how does it differ from stemming?
2. How does lemmatization handle different parts of speech?
3. Why is lemmatization considered more accurate than stemming?



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

### e) POS Tagging:

```
import nltk

# Download the averaged perceptron tagger for POS tagging
nltk.download('averaged_perceptron_tagger')

# Define the sentence
sentence = "I am learning NLP in Python"

# Tokenize the sentence
tokens = nltk.word_tokenize(sentence)

# Perform POS tagging
pos_tags = nltk.pos_tag(tokens)

# Print the POS tags
print(pos_tags)
```

### Output:

```
[('I', 'PRP'), ('am', 'VBP'), ('learning', 'VBG'), ('NLP', 'NNP'), ('in', 'IN'), ('Python', 'NNP')]
```



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

### f) spaCy Program:

```
import spacy

# Load the 'en_core_web_sm' model
nlp = spacy.load('en_core_web_sm')

# Define the sentence
sentence = "I am learning NLP in Python"

# Process the sentence using spaCy's NLP pipeline
doc = nlp(sentence)

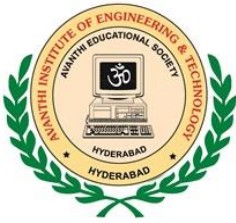
# Iterate through the tokens and print the token text and POS tag
for token in doc:
    print(token.text, token.pos_)
```

### Output:

```
I PRON
am AUX
learning VERB
NLP PROP
in ADP
Python PROP
```

### VIVA Questions :

1. What is POS tagging and why is it important in NLP?
2. Can you list the common POS tags used by NLTK?
3. How does POS tagging contribute to the understanding of a sentence's structure?



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

### g) Lexical analysis:

```
import re # for performing regex expressions

# List to hold the tokens
tokens = []

# Source code string turned into a list of words
source_code = 'int marks are given here= 100;'.split()

# Loop through each word in the source code
for word in source_code:

    # Check if the word is a datatype declaration
    if word in ['str', 'int', 'bool']:
        tokens.append(['DATATYPE', word])

    # Check if the word is an operator
    elif word in '*-/+=':
        tokens.append(['OPERATOR', word])

    # Check if the word is an identifier or contains an operator
    elif re.match("[a-zA-Z]+", word):
        # Split the word if it contains an '=' operator (e.g., 'here=')
        if '=' in word:
            parts = word.split('=')
            tokens.append(['IDENTIFIER', parts[0]])
            tokens.append(['OPERATOR', '='])
        else:
            tokens.append(['IDENTIFIER', word])

    # Check if the word is an integer and handle end statement if present
    elif re.match("^[0-9]+$", word):
        tokens.append(["INTEGER", word])
    elif re.match("^[0-9]+;$", word):
        tokens.append(["INTEGER", word[:-1]])
        tokens.append(['END_STATEMENT', ';'])
```



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanathi@gmail.com](mailto:principal.avanathi@gmail.com)

---

```
# Output the tokens  
print(tokens)
```

### Output:

```
[['DATATYPE', 'int'], ['IDENTIFIER', 'marks'], ['IDENTIFIER', 'are'], ['IDENTIFIER',  
'given'], ['IDENTIFIER', 'here'], ['OPERATOR', '='], ['INTEGER', '100'],  
['END_STATEMENT', ';']]
```

### VIVA Questions :

1. What is lexical analysis in the context of NLP?
2. How does tokenization help in text preprocessing?
3. Can you explain the difference between tokenization at the word level and at the sentence level?





# AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

## PROGRAM - 2

### 2. Sentiment analysis on customer review on products

```
import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

from nltk.sentiment.vader import SentimentIntensityAnalyzer

import nltk


# Download necessary NLTK data

nltk.download('vader_lexicon')


# Load the dataset

data = pd.read_csv("Reviews.csv")


# Inspect the first few rows and summary statistics of the dataset

print(data.head())

print(data.describe())


# Drop rows with missing values

data = data.dropna()


# Visualize the distribution of product ratings

ratings = data["Score"].value_counts()

numbers = ratings.index

quantity = ratings.values
```



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

```
custom_colors = ["skyblue", "yellowgreen", 'tomato', "blue", "red"]  
plt.figure(figsize=(10, 8))  
plt.pie(quantity, labels=numbers, colors=custom_colors, autopct='%1.1f%%', startangle=140)  
central_circle = plt.Circle((0, 0), 0.5, color='white')  
fig = plt.gcf()  
fig.gca().add_artist(central_circle)  
plt.rc('font', size=12)  
plt.title("Distribution of Amazon Product Ratings", fontsize=20)  
plt.show()
```

```
# Initialize the VADER sentiment analyzer
```

```
sentiments = SentimentIntensityAnalyzer()
```

```
# Perform sentiment analysis on the review text
```

```
data["Positive"] = [sentiments.polarity_scores(i)["pos"] for i in data["Text"]]
```

```
data["Negative"] = [sentiments.polarity_scores(i)["neg"] for i in data["Text"]]
```

```
data["Neutral"] = [sentiments.polarity_scores(i)["neu"] for i in data["Text"]]
```

```
# Inspect the first few rows of the dataset with sentiment scores
```

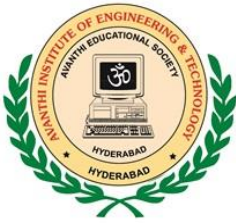
```
print(data.head())
```

```
# Calculate the total positive, negative, and neutral sentiment scores
```

```
x = sum(data["Positive"])
```

```
y = sum(data["Negative"])
```

**Avanthi Institute of Engineering and Technology**



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

```
z = sum(data["Neutral"])
```

```
# Define a function to determine the overall sentiment based on scores
```

```
def sentiment_score(a, b, c):
```

```
    if (a > b) and (a > c):
```

```
        print("Overall Sentiment: Positive")
```

```
    elif (b > a) and (b > c):
```

```
        print("Overall Sentiment: Negative")
```

```
    else:
```

```
        print("Overall Sentiment: Neutral")
```

```
# Determine and print the overall sentiment
```

```
sentiment_score(x, y, z)
```

```
print("Total Positive Sentiment Score: ", x)
```

```
print("Total Negative Sentiment Score: ", y)
```

```
print("Total Neutral Sentiment Score: ", z)
```

### Output:

The output will display:

- The first few rows of the dataset.
- Summary statistics of the dataset.
- A pie chart showing the distribution of product ratings.
- The first few rows of the dataset with added sentiment scores.
- The total positive, negative, and neutral sentiment scores.
- The overall sentiment based on these scores.



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

Id	ProductId	UserId	ProfileName
0	1 B001E4KFG0	A3SGXH7AUHU8GW	delmartian
1	2 B00813GRG4	A1D87F6ZCVE5NK	dll pa
2	3 B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres
3	4 B000UA0QIQ	A395BORC6FGVXV	Karl
4	5 B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham

	Helpfulness Numerator	Helpfulness Denominator	Score	Time \
0	1	1	5	1303862400
1	0	0	1	1346976000
2	1	1	4	1219017600
3	3	3	2	1307923200
4	0	0	5	1350777600

### Summary

### Text

- 0 Good Quality Dog Food I have bought several of the Vitality canned d...
- 1 Not as Advertised Product arrived labeled as Jumbo Salted Peanut...
- 2 "Delight" says it all This is a confection that has been around a fe...
- 3 Cough Medicine If you are looking for the secret ingredient i...
- 4 Great taffy Great taffy at a great price. There was a wid...



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

Id	ProductId	UserId ...	Positive	Negative	Neutral
0	1 B001E4KFG0	A3SGXH7AUHU8GW ...	0.305	0.000	0.695
1	2 B00813GRG4	A1D87F6ZCVE5NK ...	0.000	0.138	0.862
2	3 B000LQOCH0	ABXLMWJIXXAIN ...	0.155	0.091	0.754
3	4 B000UA0QIQ	A395BORC6FGVXV ...	0.000	0.000	1.000
4	5 B006K2ZZ7K	A1UQRSCLF8GW1T ...	0.448	0.000	0.552

Overall Sentiment: Neutral

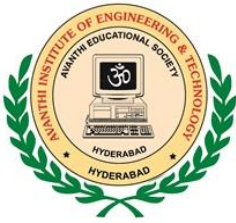
Total Positive Sentiment Score: 342.72

Total Negative Sentiment Score: 121.54

Total Neutral Sentiment Score: 535.74

VIVA Questions :

1. What is sentiment analysis, and how is it applied to customer reviews of products?
2. Why is sentiment analysis important for understanding customer reviews?
3. What are some common techniques for feature extraction in sentiment analysis?
4. How does the bag-of-words model work in the context of sentiment analysis?
5. What is TF-IDF, and how is it used in sentiment analysis?



# AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

## PROGRAM – 3

### 3. Web analytics

#### a) Web usage data (web server log data, clickstream analysis)

```
import pandas as pd
import matplotlib.pyplot as plt
import re

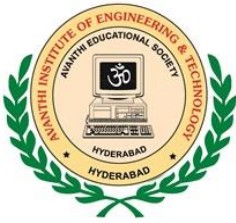
# Sample log data
log_data = """
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326
127.0.0.1 - frank [10/Oct/2000:13:56:00 -0700] "POST /form HTTP/1.1" 404 321
192.168.0.1 - jane [11/Oct/2000:14:05:36 -0700] "GET /index.html HTTP/1.0" 200 124
10.0.0.1 - bob [12/Oct/2000:15:15:36 -0700] "GET /about HTTP/1.0" 500 532
"""

# Function to parse log data
def parse_log(log):
    pattern = r'(\d+\.\d+\.\d+\.\d+) - (\w+) \[(.*?)\] "(.*?)" (\d+) (\d+)'
    log_entries = []
    for line in log.splitlines():
        match = re.match(pattern, line)
        if match:
            log_entries.append(match.groups())
    return pd.DataFrame(log_entries, columns=['IP', 'User', 'Timestamp', 'Request', 'Status', 'Size'])

# Parse the log data
df = parse_log(log_data)

# Convert data types
df['Timestamp'] = pd.to_datetime(df['Timestamp'], format='%d/%b/%Y:%H:%M:%S %z')
df['Status'] = df['Status'].astype(int)
df['Size'] = df['Size'].astype(int)

# Display the first few rows of the dataframe
```



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

```
print(df.head())

# Analyze the number of requests
request_counts = df['Request'].value_counts()

# Analyze the HTTP status codes
status_counts = df['Status'].value_counts()

# Visualize the request counts
plt.figure(figsize=(10, 5))
request_counts.plot(kind='bar', color='skyblue')
plt.title('Most Visited Pages')
plt.xlabel('Request')
plt.ylabel('Number of Requests')
plt.xticks(rotation=45)
plt.show()

# Visualize the status code distribution
plt.figure(figsize=(10, 5))
status_counts.plot(kind='pie', autopct='%1.1f%%', colors=['skyblue', 'lightgreen', 'lightcoral', 'orange'])
plt.title('HTTP Status Code Distribution')
plt.ylabel('')
plt.show()
```

### Output:

IP	User	Timestamp	Request	Status	Size
0	127.0.0.1	frank 2000-10-10 13:55:36-07:00	GET /apache_pb.gif HTTP/1.0	200	2326
1	127.0.0.1	frank 2000-10-10 13:56:00-07:00	POST /form HTTP/1.1	404	321
2	192.168.0.1	jane 2000-10-11 14:05:36-07:00	GET /index.html HTTP/1.0	200	124
3	10.0.0.1	bob 2000-10-12 15:15:36-07:00	GET /about HTTP/1.0	500	532

### VIVA Questions :

1. What information is typically contained in web server log data?
2. How can web server log data be used to analyze website performance?
3. What are some common formats of web server logs?



# AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

## b) Hyperlink Data:

```
from collections import Counter
```

```
# Sample data: list of hyperlinks
```

```
hyperlinks = [
```

```
    "https://example.com/page1",
```

```
    "https://example.com/page2",
```

```
    "https://example.com/page1",
```

```
    "https://example.com/page3",
```

```
    "https://example.com/page1",
```

```
    "https://example.com/page2",
```

```
    "https://example.org/home",
```

```
    "https://example.org/about",
```

```
    "https://example.com/page2",
```

```
    "https://example.com/page3",
```

```
]
```

```
def analyze_hyperlinks(links):
```

```
    # Count occurrences of each hyperlink
```

```
    link_counts = Counter(links)
```

```
    # Total number of hyperlinks
```

```
    total_links = len(links)
```

```
    # Number of unique hyperlinks
```

```
    unique_links = len(link_counts)
```

```
    # Most common hyperlinks
```

```
    most_common_links = link_counts.most_common()
```

```
    # Display results
```

```
    print(f"Total hyperlinks: {total_links}")
```

```
    print(f"Unique hyperlinks: {unique_links}")
```

```
    print("\nHyperlink occurrences:")
```

```
    for link, count in most_common_links:
```

```
        print(f"{link}: {count}")
```

```
    # Display the most frequent hyperlink
```

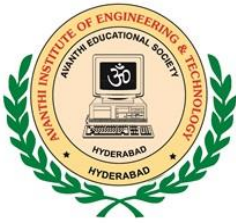
```
    if most_common_links:
```

```
        most_frequent_link, count = most_common_links[0]
```

```
        print(f"\nMost frequent hyperlink: {most_frequent_link} (occurrences: {count})")
```

Avanthi Institute of Engineering and Technology





# AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

```
# Run the analysis  
analyze_hyperlinks(hyperlinks)
```

## Output:

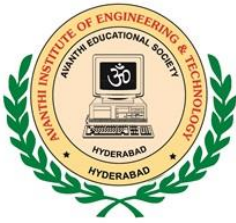
Total hyperlinks: 10  
Unique hyperlinks: 6

Hyperlink occurrences:  
<https://example.com/page1>: 3  
<https://example.com/page2>: 3  
<https://example.com/page3>: 2  
<https://example.org/home>: 1  
<https://example.org/about>: 1

Most frequent hyperlink: <https://example.com/page1> (occurrences: 3)

## VIVA Questions :

1. What is hyperlink data in web analytics, and why is it important?
2. How can hyperlink data be used to improve the user experience on a website?
3. What metrics are commonly used to analyze hyperlink data in web analytics?
4. What are the challenges associated with collecting and analyzing hyperlink data?
5. How does the analysis of outbound links contribute to understanding a website's performance?



# AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

## PROGRAM - 4

### 4. Search engine optimization- implement spamdexing

Spamdexing, also known as search engine spamming or search engine poisoning, refers to various methods used to manipulate search engine rankings to favor certain pages in ways that violate the search engine's terms of service. This is generally considered unethical and is penalized by search engines. However, for educational purposes, we can implement a basic example to understand how such techniques work.

```
def generate_spam_content(keywords, original_content, repetition=10):
```

```
    """
```

```
    Generate spam content by stuffing keywords into the original content.
```

```
    :param keywords: List of keywords to be stuffed.
```

```
    :param original_content: The original content of the page.
```

```
    :param repetition: Number of times each keyword is repeated.
```

```
    :return: Modified content with keyword stuffing.
```

```
    """
```

```
    spam_content = original_content
```

```
    keyword_block = ''.join(keywords * repetition)
```

```
    # Append keyword block to the original content
```

```
    spam_content += '\n\n' + keyword_block
```

```
    return spam_content
```

```
# Example usage
```

```
keywords = ["buy cheap products", "best prices", "discount sales", "online shopping"]
```

```
original_content = """
```

```
Welcome to our online store. We offer a wide range of products at the best prices.
```

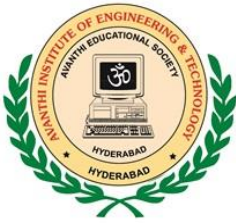
```
Browse through our collection and find the best deals for your needs.
```

```
"""
```

```
spam_content = generate_spam_content(keywords, original_content)
```

```
print("Original Content:\n")
```

```
print(original_content)
```



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

```
print("\nSpam Content:\n")
```

```
print(spam_content)
```

### Output:

Original Content:

Welcome to our online store. We offer a wide range of products at the best prices.

Browse through our collection and find the best deals for your needs.

Spam Content:

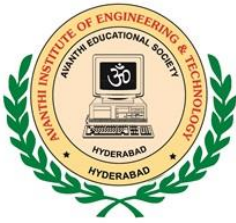
Welcome to our online store. We offer a wide range of products at the best prices.

Browse through our collection and find the best deals for your needs.

buy cheap products best prices discount sales online shopping buy cheap products best prices discount sales online shopping buy cheap products best prices discount sales online shopping buy cheap products best prices discount sales online shopping buy cheap products best prices discount sales online shopping buy cheap products best prices discount sales online shopping buy cheap products best prices discount sales online shopping buy cheap products best prices discount sales online shopping buy cheap products best prices discount sales online shopping buy cheap products best prices discount sales online shopping

### VIVA Questions :

1. What is spamdexing, and how does it affect search engine optimization (SEO)?
2. Can you describe some common techniques used in spamdexing?
3. What are the potential consequences of engaging in spamdexing for a website?
4. How do search engines detect and combat spamdexing?
5. What are some ethical SEO practices that can help avoid spamdexing?



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

### PROGRAM - 5

#### 5. Use Google analytics tools to implement the following

##### a) Conversion Statistics :

- **Google Cloud Project:** Ensure you have a Google Cloud project set up and have enabled the Google Analytics Reporting API.
- **Service Account:** Create a service account in your Google Cloud project and download the JSON key file.
- **Google Analytics View ID:** Obtain the View ID from your Google Analytics account for which you want to fetch the data.

Program:

Install Required Libraries

You'll need the google-auth, google-auth-oauthlib, google-auth-httplib2, and google-api-python-client libraries. Install them using pip:

```
pip install google-auth google-auth-oauthlib google-auth-httplib2 google-api-python-client
```

Write the Python Program:

```
import json

from google.oauth2 import service_account

from googleapiclient.discovery import build

# Path to your service account key file

KEY_FILE_LOCATION = 'path/to/your/service-account-file.json'

# Your Google Analytics view ID

VIEW_ID = 'YOUR_VIEW_ID'
```



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

```
def initialize_analyticsreporting():
```

```
    """Initializes the analytics reporting service object."""
```

```
    credentials = service_account.Credentials.from_service_account_file(
```

```
        KEY_FILE_LOCATION,  
        scopes=['https://www.googleapis.com/auth/analytics.readonly'])
```

```
    analytics = build('analyticsreporting', 'v4', credentials=credentials)
```

```
    return analytics
```

```
def get_report(analytics):
```

```
    """Queries the Analytics Reporting API V4."""
```

```
    return analytics.reports().batchGet(
```

```
        body={
```

```
            'reportRequests': [
```

```
                {
```

```
                    'viewId': VIEW_ID,
```

```
                    'dateRanges': [{'startDate': '30daysAgo', 'endDate': 'today'}],
```

```
                    'metrics': [{'expression': 'ga:goalCompletionsAll'}, {'expression':  
'ga:goalConversionRateAll'}],
```

```
                    'dimensions': [{'name': 'ga:date'}]
```

```
                }
```

```
            ]
```

```
        }
```



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC "B++" Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

```
).execute()
```

```
def print_response(response):
```

```
    """Parses and prints the Analytics Reporting API V4 response."""
```

```
    for report in response.get('reports', []):
```

```
        columnHeader = report.get('columnHeader', { })
```

```
        dimensionHeaders = columnHeader.get('dimensions', [])
```

```
        metricHeaders = columnHeader.get('metricHeader', { }).get('metricHeaderEntries', [])
```

```
        rows = report.get('data', { }).get('rows', [])
```

```
        for row in rows:
```

```
            dimensions = row.get('dimensions', [])
```

```
            dateRangeValues = row.get('metrics', [])
```

```
        for header, dimension in zip(dimensionHeaders, dimensions):
```

```
            print(f'{header}: {dimension}', end=' ')
```

```
        for i, values in enumerate(dateRangeValues):
```

```
            print(f'Values for date range {i}:')
```

```
            for metricHeader, value in zip(metricHeaders, values.get('values')):
```

```
                print(f'{metricHeader.get("name")}: {value}')
```



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

```
def main():
```

```
    analytics = initialize_analyticsreporting()
```

```
    response = get_report(analytics)
```

```
    print_response(response)
```

```
if __name__ == '__main__':
```

```
    main()
```

### Steps to Run the Program

1. **Replace KEY\_FILE\_LOCATION** with the path to your service account JSON file.
2. **Replace VIEW\_ID** with your Google Analytics view ID.
3. **Run the script:** Execute the script in your Python environment.

### Output :

The script will output the conversion statistics for the past 30 days, showing the number of goal completions and conversion rates per day.

### VIVA Questions :

1. What are conversion statistics in Google Analytics, and why are they important for businesses?
2. How do you set up and track a goal in Google Analytics to monitor conversions?
3. What is the difference between macro and micro conversions, and how can both be tracked in Google Analytics?
4. Can you explain what a conversion funnel is and how Google Analytics helps in analyzing it?



# AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

## PROGRAM - 6

### 6. Use Google analytics tools to implement the following

#### b) Visitor Profiles

```
import json
from google.oauth2 import service_account
from googleapiclient.discovery import build

# Path to your service account key file
KEY_FILE_LOCATION = 'path/to/your/service-account-file.json'

# Your Google Analytics view ID
VIEW_ID = 'YOUR_VIEW_ID'

def initialize_analyticsreporting():
    """Initializes the analytics reporting service object."""
    credentials = service_account.Credentials.from_service_account_file(
        KEY_FILE_LOCATION, scopes=['https://www.googleapis.com/auth/analytics.readonly'])
    analytics = build('analyticsreporting', 'v4', credentials=credentials)
    return analytics

def get_report(analytics):
    """Queries the Analytics Reporting API V4."""
    return analytics.reports().batchGet(
        body={
            'reportRequests': [
                {
                    'viewId': VIEW_ID,
                    'dateRanges': [{'startDate': '30daysAgo', 'endDate': 'today'}],
                    'metrics': [{'expression': 'ga:sessions'}, {'expression': 'ga:users'}],
                    'dimensions': [
                        {'name': 'ga:country'},
                        {'name': 'ga:city'},
                        {'name': 'ga:userType'},
                        {'name': 'ga:deviceCategory'},
                        {'name': 'ga:browser'},
                        {'name': 'ga:operatingSystem'},
                        {'name': 'ga:age'},
                        {'name': 'ga:gender'}
                    ]
                }
            ]
        })
```





## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

```
    ]
  }
]
}
).execute()
```

```
def print_response(response):
```

```
    """Parses and prints the Analytics Reporting API V4 response."""
```

```
    for report in response.get('reports', []):
```

```
        columnHeader = report.get('columnHeader', {})
```

```
        dimensionHeaders = columnHeader.get('dimensions', [])
```

```
        metricHeaders = columnHeader.get('metricHeader', {}).get('metricHeaderEntries', [])
```

```
        rows = report.get('data', {}).get('rows', [])
```

```
        for row in rows:
```

```
            dimensions = row.get('dimensions', [])
```

```
            dateRangeValues = row.get('metrics', [])
```

```
            for header, dimension in zip(dimensionHeaders, dimensions):
```

```
                print(f'{header}: {dimension}', end=' ')
```

```
            for i, values in enumerate(dateRangeValues):
```

```
                print(f'\nValues for date range {i}:')
```

```
                for metricHeader, value in zip(metricHeaders, values.get('values')):
```

```
                    print(f'{metricHeader.get("name")}: {value}')
```

```
            print("\n")
```

```
def main():
```

```
    analytics = initialize_analyticsreporting()
```

```
    response = get_report(analytics)
```

```
    print_response(response)
```

```
if __name__ == '__main__':
```

```
    main()
```



# AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

## Steps to Run the Program

1. **Replace KEY\_FILE\_LOCATION** with the path to your service account JSON file.
2. **Replace VIEW\_ID** with your Google Analytics view ID.
3. **Run the script:** Execute the script in your Python environment.

## Output :

The script will output visitor profile information for the past 30 days, including the number of sessions, users, and details about the visitors such as their country, city, user type, device category, browser, operating system, age, and gender.

## VIVA Questions :

1. What is a visitor profile in Google Analytics, and why is it important for understanding website traffic?
2. How can you use Google Analytics to segment visitors based on demographics and interests?
3. What is the importance of using custom segments in Google Analytics, and how do you create one?



# AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

## PROGRAM - 7

### 7. Use Google analytics tools to implement the Traffic Sources.

Write the Python Program

Here is a Python script to authenticate and retrieve traffic sources information from Google Analytics:

```
import json

from google.oauth2 import service_account

from googleapiclient.discovery import build

# Path to your service account key file

KEY_FILE_LOCATION = 'path/to/your/service-account-file.json'

# Your Google Analytics view ID

VIEW_ID = 'YOUR_VIEW_ID'

def initialize_analyticsreporting():

    """Initializes the analytics reporting service object."""

    credentials = service_account.Credentials.from_service_account_file(

        KEY_FILE_LOCATION, scopes=['https://www.googleapis.com/auth/analytics.readonly'])

    analytics = build('analyticsreporting', 'v4', credentials=credentials)

    return analytics
```



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

```
def get_report(analytics):
```

```
    """Queries the Analytics Reporting API V4."""
```

```
    return analytics.reports().batchGet(
```

```
        body={
```

```
            'reportRequests': [
```

```
                {
```

```
                    'viewId': VIEW_ID,
```

```
                    'dateRanges': [{ 'startDate': '30daysAgo', 'endDate': 'today' }],
```

```
                    'metrics': [{ 'expression': 'ga:sessions' }, { 'expression': 'ga:users' }],
```

```
                    'dimensions': [
```

```
                        { 'name': 'ga:source' },
```

```
                        { 'name': 'ga:medium' },
```

```
                        { 'name': 'ga:campaign' }
```

```
                    ]
```

```
                }
```

```
            ]
```

```
        }
```

```
    ).execute()
```

```
def print_response(response):
```



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

```
"""Parses and prints the Analytics Reporting API V4 response."""
```

```
for report in response.get('reports', []):
```

```
    columnHeader = report.get('columnHeader', { })
```

```
    dimensionHeaders = columnHeader.get('dimensions', [])
```

```
    metricHeaders = columnHeader.get('metricHeader', { }).get('metricHeaderEntries', [])
```

```
    rows = report.get('data', { }).get('rows', [])
```

```
    for row in rows:
```

```
        dimensions = row.get('dimensions', [])
```

```
        dateRangeValues = row.get('metrics', [])
```

```
        for header, dimension in zip(dimensionHeaders, dimensions):
```

```
            print(f'{header}: {dimension}', end=' ')
```

```
        for i, values in enumerate(dateRangeValues):
```

```
            print(f'\nValues for date range {i}:')
```

```
            for metricHeader, value in zip(metricHeaders, values.get('values')):
```

```
                print(f'{metricHeader.get("name")}: {value}')
```

```
        print('\n')
```

```
def main():
```



## AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, Recg. By Govt. of T.S & Affiliated to JNTUH, Hyderabad)

NAAC “B++” Accredited Institute

Gunthapally (V), Abdullapurmet(M), RR Dist, Near Ramoji Film City, Hyderabad -501512.

[www.aietg.ac.in](http://www.aietg.ac.in) email: [principal.avanthi@gmail.com](mailto:principal.avanthi@gmail.com)

---

```
analytics = initialize_analyticsreporting()
```

```
response = get_report(analytics)
```

```
print_response(response)
```

```
if __name__ == '__main__':
```

```
    main()
```

### Steps to Run the Program :

1. **Replace KEY\_FILE\_LOCATION** with the path to your service account JSON file.
2. **Replace VIEW\_ID** with your Google Analytics view ID.
3. **Run the script:** Execute the script in your Python environment.

### Output :

The script will output traffic sources information for the past 30 days, including the number of sessions, users, and details about the traffic sources such as source, medium, and campaign.

### VIVA Questions :

- a. What are traffic sources in Google Analytics, and why are they important for understanding website performance?
- b. How do you categorize traffic sources in Google Analytics, and what are the main categories?
- c. How can you set up and track UTM parameters to measure the effectiveness of specific traffic sources?