Infosys Springboard
Virtual Internship 6.0
AirAware
Project Document

Name: Jampana Rahul Varma

Batch: 6

Team: 1

Email: rahulvarmajampana1302@gmail.com

Day 1 – Overview of Project:

Topic - AirAware Smart Air Quality Prediction System

AirAware is a smart system designed to monitor and predict air quality using data analysis and machine learning.

Real-time air pollution data containing PM2.5, CO_2, NO_2, etc.

Selected Frontend, Backend, Database, Dataset.

Air Quality Monitor using ML models and Data Analysis

Real time Air pollution data from websites like Kaggle or HuggingFace

NO2 CO2 O2 levels are some columns to appear in dataset

Forecast future air quality and give awareness

Visual analysis

Project for next 45 days

Objective:

- Collect dataset with PM2.5 CO2 NO2 etc.
- Predicted vs Actual Air quality
- 3 to 4 months data and analytical part should be in dashboard
- Alert user if air quality becomes hazardous
- Get weather from sensors or dummy input and store it in database
- Get dataset from Kaggle and preprocess it by normalizing and cleaning it
- Then feed it to a model like Linear Regression, Random Forest or SVM
- Frontend language Backend language Database used to store the data and from where you are collecting the dataset and complete overview of the project
- Short and brief intro about your project like what it is and how you are going to approach it

4 milestones:

- 1st milestone - 25%
- 2nd milestone - 50%
- 3rd milestone - 75%
- 4th milestone - 100%

Day 2 – Team Allocation and basic instructions:

Team 1 - Rajalaksmi, Rahul, Sreya, Lokesh, Divija Nandana - Air quality data of Delhi, India Kaggle

Instructions:

- Predict the Quality of Air - Is it going to be more polluted in 4 or 5 years if so how to avoid
- Show all simple particular components like Last year air quality, Present air quality, predictions etc. in UI dashboard
- Different UI for each team
- Use API calls to communicate with backend
- Send user UI interaction as payload to backend code which is then processed by backend and sent as payload to frontend
- Graph of last few years
- CO2,PM,N2 levels if levels are in certain range pure air else unpure air
- Currently dummy data for UI presentation
- Try to achieve 95% to 100% accuracy rate
- Accuracy rate should be displayed using heatmap on dashboard
- Different blocks in dashboard
- Flask or FastAPI for backend
- Dataset should be in database and displayed in dashboard with dynamic updates

Future Directions:

- Current weather status
- Google weather prediction integration
- Air quality app with sensors to check surrounding air quality could be developed in future if your app is completed quickly
- Input manually from your end
- Future enterprise level project uses sensor data
- Application of enterprise level for air quality prediction

Presentation of codes from Monday

Day 3 - APIs:

Flask or FastAPI are main APIs used in Python

API:

1. Fast API
2. Flask API

Fast API:

```
from fastapi import FastAPI

app = FastAPI()

@app.get("/")
```

Flask:

```
from flask_api import FlaskAPI

from flask import request

app = FlaskAPI(__name__)

@app.route("/", methods=['GET'])
```

Method is declared inside parenthesis for Flask and outside parenthesis for FastAPI.

postman is a testing tool used to test the backend API without UI.

Commonly used methods are GET and POST.

In GET payload is sent in header. Used to retrieve data from frontend.

In POST payload is sent in body. If we want to give a response to frontend we use POST method.

Payload is the input given to frontend.

Backend should have the ability to answer to the payload.

Response is the information sent by backend to frontend.

Payload+Response processed is shown in UI.

API is used for communication between frontend and backend.

While designing API you need to know how frontend is going to send payload(payload structure). Otherwise 402 error and frontend will not give payload to backend and API will not run.

Get payload structure from frontend before writing backend.

If payload structure changes code will not work.

Flask is more comfortable for users. FastAPI is advanced compared to Flask. FastAPI is built using pydening used for building fast apps and has async function support. Flask has only sync function. Async is used for multiple endpoints. FastAPI is used for high concurrency.

Synchronized processing is sequence processing. Asynchronized is parallel processing.

Django uses REST framework. Enhanced version of REST framework is Flask. In Flask we wrap the functions. Flask is slower, easier and simpler than FastAPI.

Websocket - A WebSocket is a communication protocol that provides a persistent, full-duplex (two-way) connection between a client (browser/app) and a server.

WebSocket was used before HTTPS. In WebSocket we use ws it is a little old ws(HTTP) and wss(WebSocket Security)(HTTPS). Sends response continuously to UI. Steaming response backend will continuously interact with frontend to give responses Flashing response backend sends response to frontend at once.

Streaming - Streaming refers to sending data in chunks, continuously, instead of sending everything at once.

What is API? What are the types of APIs we discussed today?

Try to build a basic project on AirAware. Build a beautiful eye warming UI dashboard.

Day 4 - Git:

If more than one person works on a project Git is used to store the project. One or more files can be stored. GitHub repository is shared.

Commands:

- git add . - Used to add all newly created files in the current folder to a repo
- git add main.py - Used to add only particular file to folder
- git commit -m "Message" (like fhdhl) - Creates commit id which is unique with commit message
- git push - Used to push changes to repo. Only after you push the changes will be visible
- git branch - Shows current feature branch you are working on
- git branch --all - Shows all feature branches that were created so far for the project
- git fetch --all - Gives updated changes from all branches
- git pull - Used to get changes pushed by others using the same branch. Gets to upto date status
- git stash - Used to save changes to local without pushing changes to repo
- git checkout -b "sh_new" -
- git checkout "sh_new" - Switches from current branch to branch inside quotations

Virtual Environment:

- python -m venv venv - Used to create a virtual environment in python
- venv/Script/Activate - Used to activate virtual environment
- requirements.txt - Has a list of all the libraries that are required to be installed for the project to work
- pip install -r requirements.txt - Used to install all libraries in requirements.txt using pip
- deactivate - Used to come out of the environment we are currently in

Ctrl+J - Opens Terminal in VSCode

Day 5 - Database:

DB - Database

DBMS - Database Management System

CRUD - Create Read Update Delete

Structured - MySQL, SQL, PostgreSQL

Unstructured - MongoDB

Normalization - 1NF, 2NF, 3NF, BCNF, 4NF, 5NF

1NF - First Normal Form

2NF - Second Normal Form

3NF - Third Normal Form

BCNF - Boyce Codd Normal Form

4NF - Fourth Normal Form

5NF - Fifth Normal Form

Partial Dependency - A partial dependency occurs when a non-prime attribute (non-key attribute) depends on part of a composite primary key, not the whole key.

Transitive Dependency - A transitive dependency occurs when a non-prime attribute depends on another non-prime attribute, not directly on the primary key.

Primary Key - A primary key is a unique identifier for each record in a table.

Foreign Key - A foreign key is an attribute in one table that refers to the primary key of another table.

Composite Key - A composite key is a primary key made of two or more columns together to uniquely identify a row.

Master Data is fixed it will not change

Transaction Data will change with time

Day 6 – AI/ML:

Providers:

1. Google
2. OpenAI
3. Grok

Models:

1. Gemini
2. GPT
3. Llama

LLM - Large Language Model

Libraries:

1. openai - GPT
2. genai - Gemini
3. grok - Grok, Llama

Platforms:

1. Google AI Studio - Gemini
2. OpenAI - GPT
3. Grok - Grok

API key has rate limit and token limit for usage

Free trial is also available for some models with limited model calls

Chat completion component - Has parameters like max_tokens etc.

Control of the model is in the prompt

Prompt Engineer - Trains the model based on prompting

Prompt Engineering - Training the model using prompts

ML:

Types:-

1. Supervised
2. Unsupervised
3. Reinforcement Learning

Supervised:-

- Labeled data
- Regression and Classification
- Linear and Logistic Regression
- Classification is done based on labels

Unsupervised:-

- Unlabeled data
- Clustering and Association Mining
- K-Means Clustering
- K is the number of clusters to be formed

Reinforcement Learning:-

- Learning from mistakes

Postgres is done using pgAdmin application. Postgres is mainly used to store vectors.

Vector is a numerical representation of text with dimensions. 1/36 is highest dimension vectors.

Embedding - Conversion of text to vectors

NLTK - Natural Language Toolkit. Library used for embedding.

Day 7 – Doubt clarification:

Cloning a repository on system:

step 1 - download the git in local system.

step 2 - copy the github url you created(link).

step 3 - open the command prompt or terminal and type the "git clone"

Presentation should have code, algorithm, logic, libraries

Team 1 – 20th November 2025

Personal GitHub repository link - https://github.com/jrv1302/AirAware

Team GitHub repository link - https://github.com/SaiSreya96/grp1_team_repo_infosys

Day 8 – Milestone 1 preparation:

For Milestone 1, we were given time to focus on preparing the necessary project documentation and ensuring that all foundational components were clearly defined. This included identifying the project objectives, and finalizing the functional and non-functional requirements. The document was organized in a structured format to maintain clarity and to align with the guidelines provided for Milestone 1.

Day 9 – Milestone 1(Day 1):

During Milestone 1, we completed the foundational planning, documentation, and initial implementation of the AirAware – Smart AQI Prediction System. All conceptual and design-related work discussed up to 20 November 2025 was compiled into a structured project document and presented to the review panel. Presented the project idea, motivation (Delhi pollution crisis), and real-world relevance. Milestone 1 successfully established the UI skeleton, computational logic base, and project direction.

Day 10 – Milestone 1(Day 2):

Completed milestone 1 with all teams presenting their progress and getting evaluated by the instructor.

Day 11 – Artificial Intelligence:

What is Artificial Intelligence?

Artificial intelligence (AI) refers to the capability of machines to perform tasks that typically require human intelligence, such as learning, reasoning, and problem-solving

Key subfields:-

ML, DL, NLP, RL, Knowledge-based sytems

- ML - Algorithms that learn from the data. Supervised, unsupervised and reinforcement learning.
- DL - Uses Neural Network with many layers to learn patterns. NN, ANN, CNN, RNN, Transformers.
- NLP - Natural Language Processing
- RL - Reinforcement Learning. Agents.
- KBS - Knowledge based systems are rule based systems. They have rules to restrict them.

DL Models:-

ANN, RNN, CNN, LSTM, Transformers

1. ANN - Artificial Neural Network
2. CNN - Convolutional Neural Network
3. RNN - Recurrent Neural Network
4. LSTM - Long Short Term Memory(Special type of RNN). Used to remember long term information, reduces vanishing gradient, avoid forgetting patterns, good for time-series data. Has three gates - forget gate, input gate, output gate.

   Forget gate - Decides what old information to forget

   Input gate - Decides what information to take

   Output gate - Decides what information to pass to next cell

How to work on ML:-

Step 1: Collect data

Step 2: Prepare & engineer features(Preprocessing step)

Step 3: Train a model

Step 4: Evaluate on holdout data using metrics

LLMs are designed based on DL techniques.

Day 12 - NLP:

Natural Language Processing

Subset of AI

Teaching computer to read understand and talk like a humans

Helps computer to understand, interpret and generate human languages

Gap between the human and machine Meanings, grammar, context, emotions, ambiguity

No search engines, chatbots, apps without NLP

Helps machine to understand text, speech and customer feedbacks

Applications:-

Chatbots

Semantic Analysis:-

Positive

Negative

Neutral(rarely used)

NLTK:-

Basic logic of NLP is implemented using NLTK(Natural Language Toolkit) library in python

pip install nltk

import nltk

nltk.download('all')

NLTK has 7 features:

1. Tokenization - Breaking text into words
   Eg: NLP is interesting
   ["NLP", "is", "interesting"]

   from nltk.tokenize import word_tokenize

2. Stop word removal - Words that don't have meaning come under stop words. Eg: is, the, and, or, are

   from nltk.corpus import stopwords

   Eg: this is a sample sentence
   this sample sentence

3. Stemming - Converts word to base form(Gives root word). Eg: playing --> play, studies --> studi, better --> bet

   from nltk.stem import PorterStemmer

4. Lemmatization - Converts word to base form using dictionary rule(based on meaning of word). Eg: better --> good

   from nltk.stem import WordNetLemmatizer

5. Part of Speech(POS) Tagging - Giving a tag to a word in a sentence.

   from nltk.tokenize import word_tokenize

6. Named Entity Recognization - Recognizes names, places, dates, organizations etc. from sentences.

   nltk.download('words')
   nltk.download('maxent_ne_chunker')

7. Text Preprocessing - Sentences and words will be converted to lower case. Punctuation like exclamations, quotations are removed. Removing the numbers. Removing extra spaces. Tokenization. Stop-word removal. Lemmatization or stemming.

Content is ready to train the nlp models.

Convert raw text to preprocessed text and give to model for training. These techniques are used to remove noise in text. In this case stop words, continuous tense etc. which are unwanted to machines are considered noise.

Models - SVM, Logistic Regression, TF-IDF, bag of words

Logistic Regression is a classification algorithm. Can be used in spam detection, sentiment analysis.

SVM - Support Vector Machine. It is one of the most powerful ML algorithms. Finds boundaries using linear or non-linear SVM that separates classes with hyperplane.

TF-IDF - Term Frequency - Inverse Document Frequency. TF calculates how often a word appears in a document. IDF calculates how rare the word is across all documents.

Eg: 1 out of 1000 documents --> high IDF. 999 out of 1000 documents --> low IDF.

Day 13 – SVM and RL:

SVM:

Support Vector Machine

Comes under supervised learning

Can be used for Classification, Regression, outlier detection

Mainly used for classification

Margin - Distance between the separating line and the closest data points from each class.

Closest point is called support vector

The closest point == support vectors

Hard margin, Soft margin

Hard Margin - Perfect separation. No misclassification. Data is clean. No noise. Perfectly separable. But it will not exist in real time.

Soft Margin - C parameter. Non-linear separated curve. Data will be misclassified. Used in real life.

RL:

Reinforcement Learning

Base for Agentic AI

An agent learns by interacting with an environment and receiving rewards

Nvidia, Google, Microsoft are building agents and releasing but they are not production level yet.

Agent will have capability to take decisions based on predefined datasets. It is a replica of human intelligence.

Agent with Tools

Tools are like a function

If there is only one agent then it is called single agent architecture.

If there is more than one agent then it is called multi-agent architecture. One manager agent with different agents.

Agent will take a decision by its own to select the tools. The agent will decide which tool is needed and which tool will give output for the question. The agent will take decision based on the characteristics of the tool.

Types of agents:-

1. Human-in-loop
2. Without human intervention

Human-in-loop - If you ask any question to the agent, the agent will ask you shall we continue with the decision.

Without human intervention - It will execute decision and if it is wrong then program fails only then we will know.

Agents are in development stage so we use human-in-loop agents.

Autogen is a framework used in python. It is also available as Autogen AI studio with UI. It is production level ready platform. Nowadays to build a python project you can use agentic framework.

Agentic platforms - QAI, Autogen.

Agents in autogen - User proxy agent, assistant agent etc.

Weather prediction agent etc. could be used in project.

Master agent - manager agent.

Types of agents:-

1. Simple Reflex
2. Model Based
3. Goal Based
4. Utility Based

Parameters in agents like never, medium etc. are used to determine how much human will be involved in the loop.

Day 14 – CSS Sizing:

CSS Sizing:

Intrinsic Size == orginal size

Extrinsic Size == we will use the class atributes we will give the size over the image

Handling overflow:-

overflow:

visible

hidden

scroll

auto

we have overflow-x and overflow -y

Min and Max size

Meta Element - Additional element

Box Sizing

content - margin, padding, border

content box and border box

universal selector

layout

flexbox

css grid

flex attributes

a. display:-

1. flex
2. inline-flex
3. grid
4. none

b. flex direction

justify-content: center

align-items

HTML & CSS
HTML = Using this we create the webpage
CSS = Styling elements

HTML Tags <h> <p><img>

<header> this is the header and top layer
<nav> navigation url will be here
<main> body of the content in the webpage
<footer> is the bottom content

<video>, <audio>

color, alignment

flexbox makes the one dimentional layout

css grid two dimenstional layout

python integration to html

@app.route('/')

def index ():

return render_template('index.html')

Git - Version control

GitHub - Cloud platform to host git changes

Day 15 – Flexbox and AI Models:

Milestone 2 - 4th and 5th December

Submit document and create a simple chatbot using OpenAI API or Gemini API in individual repository

Push codes to group repository with code having some functionality

Flexbox:

Flex wrap-used to adjust
Now wrap(default)
Wrap
Wrap reverse(things are arranged in reverse order)

 bg container (flex container)
Direct children will become flex items
Cards container also becomes a flex container

Align self
Flex start
Centre
Flex end
Stretch
Auto(default)

flex-wrap

nowrap
wrap
wrap-reverse
bg contqiner
Align Self
flex-start
center
flex-end
stretch
auto

Order
1.0(default)comes first
2.+ve
3.-ve (come after 0's)

pip install openai

```
from openai import OpenAI

client = OpenAI()

resp = client.chat.completions.create(

      model="gpt-4o-mini",  # replace with an available model like "gpt-4o-mini" or "gpt-4o" or "gpt-5" etc.

      messages=[

         {"role": "system", "content": "You are a helpful assistant."},

         {"role": "user", "content": "Write a two-line poem about chai."}

      ],

      max_tokens=120

   )
```

model
message
1.system
2.user
3.max_token

temperature - Affects creativity of the model. 0.1 to 1.0. 0.1 - no creativity. 0.5 to 1.0 - creative.

Day 16 – ML Algorithms:

Milestone - 2 Thursday and Friday (4th and 5th Dec)

Groupwise presentation and individual presentation, Develop a simple chatbot with chat completion, ppt, documentation, no ppt for chatbot

What actually chat completion is, how it works could be added to the document

ML Models:

Logistic Regression:-

Logistic Regression is used for classification.

$z=w1x1+w2x2+...+b$

$p=\sigma(z)=1/1+e-z1$

$p>0.5$ - class 1

else class 0

```
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
```

Decision Tree:-

Splits data into smaller groups based on questions.

shall i go out and play for a while

sunny or rainy

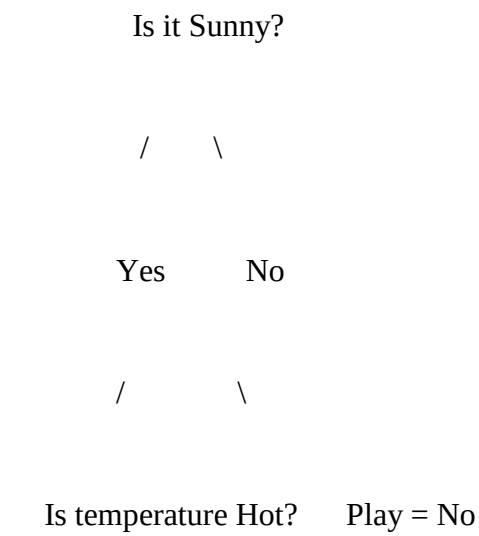
rainy you should not play

sunny

it is so hot or normal

hot then you should not play

if it is normal you can play



Information Gain(IG) and Entropy are used to determine the tree structure

Entropy=−plogp−qlogq

IG=Entropy(parent)−Entropy(children)

```
from sklearn.tree import DecisiontreeClassifier
clf = DecisionTreeClassifier()
```

Random Forest:-

Answers based on multiple decision trees. Can perform both classification and regression. If decision trees close into even numbers then classification. If decision trees close into odd numbers then regression.

Final output = majority vote (classification) or average (regression).

Builds multiple decision trees, each seeing a random part of data.

```
from sklearn.ensemble import RandomForestClasifier
rf = RandomForestClassifier()
```

K-Nearest Neighbors (KNN):-

d=sqrt((x1−x2)^2+(y1−y2)^2)

```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors =5)
```

Eg:

A = 170

b = 168

c= 150

d = 152

a b is good for sports based on there height

c and d or not fit.

169

151

K Means Clustering:-

It comes under unsupervised learning.

Groups data into K clusters based on similarity.

Steps:

- Choose K centroids
- Assign points to nearest centroid
- Recalculate centroid
- Repeat

$\sum||xi-centroid||^2$

from sklearn.cluster import KMeans

kmeans = KMeans(n_clusters =3)

Basic ML algorithms come under sklearn

Linear Regression:-

Monitors and predicts continuous values and elaborates the basic linear algorithm which follows the line y = mx+c

Predicts continuous values (sales, price, temperature).

y-mx+c

MSE=1/n∑(y−y^)^2

from sklearn.linear_model import LinearRegression

lr = LinearRegression()

XGBoost:-

Sequential trees will make mistakes so next model will correct mistakes of previous models.

Fnew(x)=Fold(x)+η·h(x)

Where:

η = learning rate

h(x) = small decision tree

XGBoost adds regularization

Gives very high accuracy but takes more time.

xgboost is separate package. XGBClassifier.

from xgboost import XGBClassifier

model = XGBClassifier()

Day 17 - MySQL:

SQL = Structured Query Language

SELECT * FROM table name;

Some of SQL Commands:-

SELECT - Extracts data from database

Update - Updates the database

Delete - Deletes data from the database

Insert - Inserts data into the database

Create -

ALTER - Used to modify the database

DROP -

RENAME -

truncate -

select column1 , column2 from tablename - Command to get only two columns from database

SELECT student name, student id FROM student;

SELECT * - Shows the entire table

SELECT column - Shows only the column

SELECT DISTINCT - Used to select distinct values from a column

WHERE - Used to select anything particularly

SELECT * From Customer WHERE country = 'Mexico';

Operators:

=

>

<

>=

<=

<> or !=

BETWEEN

LIKE

IN

AND

OR

NOT

ORDER BY(Default is ASC) ASC or DESC

Eg:

select * from students where student id=1;

select * from employee where age>=21

SELECT *

FROM Employees

WHERE Department = 'Sales' AND Salary > 50000;

select * from student where cgpa>9 and attendance> 90

SELECT * FROM Employee

WHERE Experience>5 AND Salary>10000 OR Salary<100000

select * from student where name like kapil and marks = 100 or name like Raju

SELECT column1, column2

FROM table_name

WHERE condition1 AND (condition2 OR condition3);

SELECT * FROM employees WHERE (name = 'Alex' OR name = 'David') AND id > 1;

INSERT INTO - To insert values into table INSERT INTO tablename col1, col2, col3..
VALEUS val1, val2

Update - UPDATE tablename SET col = val WHERE Conditon

SUM

MIN

MAX

COUNT

LIKE:-

Operators:

a%

%a

%a%

_a%

a_%

a__%

a%o

Eg:

Apple

Banana

Ball

Ram

Arc

Army

AIO

IN - SELECT * FROM employees WHERE name IN ("rahul", "kamal");

NOT IN - SELECT * FROM employees WHERE name NOT IN ("Rajesh", "Ajay")

BETWEEN - SELECT * FROM employees WHERE Salary BETWEEN 10000 AND 20000;

ALIASES

JOINS:

INNER JOIN

CROSS JOIN

LEFT JOIN

RIGHT JOIN

INNER Join:-

Table 1:

std id, std name, address

table 2:

sbj id, stud id, sbj name

result:

stud id

CROSS Join:-

Table 1:

std id, std name, address

table 2:

sbj id, stud id, sbj name

result:

stud id, std name, address, sbj id, sbj name

LEFT Join:-

Table 1:

std id, std name, address

table 2:

sbj id, stud id, sbj name

result:

std id, std name, address, stud id

RIGHT Join:-

Table 1:

std id, std name, address

table 2:

sbj id, stud id, sbj name

result:

stud id, sbj id, stud id, sbj name

UNION - Only distinct values

UNION ALL - All values

UNION ALL contains duplicates whereas UNION are unique

GROUP BY - To group the columns based on the aggregate functions. Group by category

HAVING - WHERE in Aggregate functions

Day 18 – Preparation for Milestone 2:

We prepared our document, ppt and code for milestone 2.

Document - classes

PPT

code for the project

open api - Using code

you can use any model

Day 19 – Milestone 2 Day 1:

Some of the students have presented their progress in the project and received feedback from the instructor.

Day 20 – Milestone 2 Day 2:

Remaining students have presented their progress and received feedback from the instructor.

Day 21 – Project Presentation Discussion:

The instructor has provided the required information for Monday's team presentation.

Day 22 – Inbuilt methods in Python:

Inbuilt string methods in python:

lower - Converts to lower case

upper - Converts to upper case

split - Splits words one by one

strip - Removes white spaces

join - Joins words

replace - Replaces parts of string with given string

startswith - Checks if string starts with a given string

endswith - Checks if string starts with a given string

find - Finds if a given string occurs in a string

isdigit - Checks if string is numerical

isalpha - Checks if string has alphabets

Day 23 – Inbuilt methods in Python continuation:

List Methods:

append()

extend()

insert()

remove()

pop()

index()

count()

sort()

reverse()

pop() - Remove key and returns value

popitem() - removes last inserted key-value pair

Set Methods:

add()

remove()

discard()

pop()

clear()

union()

intersection()

difference()

remove() - removes an element.

discard() - Removes an elemet if present

pop()- removes and returns an elemets.

clear() - removes all the elements

a = a,b,c

b = c,d,e

union = a,b,c,d,e

intersection = c

diference = a-b = a,b

difference = b-a = d,e

file i/o Methods:

open()

read()

readline()

readlines()

write()

writelines()

close()

General Purpose:

len()

range()

print()

type()

id()

sorted()

enumerate()

zip()

Day 24 – Inbuilt methods in Python continuation:

Conversion Functions:

int()

float()

str()

list()

dict()

set()

tuple()

Mathematical Functions:

abs()

sum()

min()

max()

pow()

round()

Functional programming tools:

filter

map

reduce

Parameters

Attributes

lambda

```
a = lambda a,b : a*b
print(a(3,4))
```

Input and Output methods:

input()

format()

class and object related:

getattr()

setattr()

hasattr()

delattr()

isinstance()

issubclass()

Object - it is the instance of some class, type or the subclasses

Miscellaneous:

globals()

locals()

collable()

eval()

exec()

Exection Handling:

try

except

finally

Memory and object Management:

del()

gc.collect()

collect() - Forces garbage collection

import gc

Simple code for gc for milestone-3

Working with iterables:

next()

iter()

decorators and Metaprograming:

stacticmethod()

classmethod()

context Managers:

with/as

import()

Day 25 – Python Interview Questions:

Python Interview Questions:

What are python's key features?

Interpreted

High level

Dynamically typed

FOSS

Object oriented

Cross Platform

Extensive Libraries are available

what are Python's data types?

int

float

complex

string

list

tuple

set

dict

Boolean

None

What is PEP 8 and why is it important?

Style pattern of Python used for structured code

What are Mutable and immutable data structure in Python?

String, tuple are immutable others are mutable

What are Built in data structures in Python?

list, tuple, set, dict

What are python memory management features?

Private heap, reference counting, garbage collector

Explain indentation in Python.

Defines code block of function

How is Python interpreted?

Compiled into byte code

Explain namespaces.

Mapping namespaces to object

What is the difference between list and tuple?

List mutable, Tuple is immutable. Tuple is faster than List.

How are sets used?

Stores unique elements faster

What is dictionary?

Key value pair

How to merge dictionaries?

Using update() method

dict1.update(dict2)

How to remove duplicates from list?

Convert to set.

What is flattened nested list?

More than 1 list will be in a list

What is the difference between shallow and deep copy?

Shallow copy - Copy of reference

Deep copy - Clone of objects

How slicing works?

Using start, end and step like s[0:5:2]

How to define reverse list?

s[::-1]

What is frozen set?

Immutable set

What is the difference between is and ==?

is gives True or False by checking data type

== is mathematical operation that checks values

Implement stack and queue.

```
from collections imports deque
q = deque([])
q.append("A")
q.append("B")
q.popleft()
stack = []
stack.append("A")
stack.append("B")
stack.pop()
```

In corporate gc with methods

Day 26 – Python and Java Interview Questions:

Thursday and Friday Milestone-3

Python Interview Questions:

What is functions vs methods?

Function is independent

Methods are bounded to objects

What are args and kwargs?

args - variable positional

kwargs - variable keywords

What are Recursion functions?

They call themselves multiple times

What are decorators?

Extend the behaviour of a particular function

What are generators?

They are memory efficient functions that yield a value

What is the difference between iterator and generator?

iter() and next() methods are used in iterators

Generators auto create them with yield

What is static method and class method?

Static method is same for all objects

Class method is

What is monkey patching?

It dynamically changes class behaviour at runtime based on user input(It is a new thing)

What is Error vs Exception?

Error is irrecoverable. It occurs during Runtime.

Exception can be handled. Handled using try except.

What is a custom exception?

Subclass of exception that can be used to create new exceptions

What is exception chaining?

One exception will be caused by another exception

What will happen to program if exception is not handled?

Program will get terminated

What is model vs package?

Model is a single py file called in code

Package is a collection of py files

What does pip stand for?

Python installer package

How to create a python virtual environment?

python -m venv venv

Activate

What is the difference between tell and seek?

Tell is a current pointer

Seek is a move pointer

How to check if a file exists?

os.path.exists(filename)

What is the difference between absolute and relative inputs?

Absolute will install all modules in a package

Relative will only install required modules

shallow.copy() - To copy file

shallow.move() - To move file

What is pivot in Pandas?

It will reshape the data into tabular format

What is groupby in Pandas?

Splits data, applies function and returns result

What is concat vs append?

Append - Add a single element to the end of an existing list in-place.

Concat - Combine multiple data structures (lists, DataFrames, Series) along a specific axis (row-wise or column-wise).

What is GIL(Global Interpreter Lock)?

The Global Interpreter Lock (GIL) is a mutex (mutual-exclusion lock) in CPython (the standard Python interpreter) that ensures only one thread executes Python bytecode at a time, even on multi-core processors.

What is meta class?

Classes of class control class creation is called meta classes

What is multi threading?

Run multiple threads at a time

What is multi processing?

Runs multiple processes at a time

What is feature scaling?

Standardization, Normalization to equalize features

How to split train and test data using a single function?

train_test_split

Java Interview Questions:

What are the features of Java?

Object oriented
High level language
Robust
Multithreaded
Platform independent

What does JDK stand for?

Java Development Kit

What does JRE stand for?

Java Runtime Environment

What does JVM stand for?

Java Virtual Machine

What is the difference between JDK JRE JVM?
JDK has tools to run Java programs
JRE is the runtime
JVM executes Bytecode

What are access modifiers in Java?

private
public
protected
default

What is the difference between == and .equals()?

== refers to value
.equals() refers to memory

What is constructor?

Used to instantiate an object

Can constructor be private?

Yes

What is difference between method overloading and method overriding?

Multiple methods with same name but different parameters
Sub class provides a specific implementation of a method from parent class

What are primitive data types?
int, float, char, byte

What is type casting?
Converting one type to another

Difference between static and non-static variables?

Static belong to a class

Non-static belong to an instance

Difference between final, finally and finalize()?

final makes a variable constant

finally is a block that is always executed

finalize is a method that is called before garbage collection

What is the use of this and super?

this is used to refer to current class object

super refers to parent classes

What is the difference between String, StringBuffer and StringBuilder?

String is immutable(for security reasons, synchronization and caching)

StringBuffer and StringBuilder are mutable

What is the difference between Array and ArrayList?

Array has fixed size

ArrayList has variable size

What is collection framework?

It is a set of classes and interfaces for storing and manipulating group of data

What is list vs set?

List is unordered and allows duplicates but set does not

What is ArrayList vs LinkedList?

ArrayList is faster in access

LinkedList is better for insertion and deletion

What is concurrent HashMap?

It is a thread safe version of HashMap without locking entire map

What is Failfast vs Failsafe?

Failfast throws concurrent modification exception

Failsafe does not

Can we have multiple catch blocks?

Yes. If we handle different exceptions

What is the difference between throw and throws?

throw is for Unchecked exception

throws is for Checked exception

What is the difference between process and thread?

A process is a self-contained execution environment with its own memory space, while a thread is a lightweight unit of execution that shares memory and resources with other threads within the same process.

What is synchronization?

Prevent thread interface by locking shared resources

What is difference between wait and sleep?

In sleep thread will not release lock

In wait thread will release lock

What is deadlock?

If two or more thread will wait for each other then it is called deadlock

What is stream API?

It is used for processing collections of objects in a functional, declarative way.

What is serialization in Java?

Converting an object into a byte stream for storing is known as serialization

What is inner class?

A class declared within another class is known as inner class

What is the difference between compile time and runtime polymorphism?

Compile time Method overloading

Runtime Method overriding

What is method hiding?

Static method in sub classes with same signature as in super class

Day 27 – Preparation for milestone-3:

Prepared for milestone 3 by coordinating with team members.

Day 28 - Software Development Lifecycle:

Teams in SDLC:-

Functional consultant/Business Analyst - They will get requirements from client and prepare a document called functional specification document. Based on that document they will configure some changes.

Developer - They will go through the functional specification document and understand client requirements. They will rewrite the functional specification document into technical specification document. Based on the technical specification document they will design a prototype to get approval from functional people. Then they will do the project. Once project is completed they will perform Unit Testing i.e. test every scenario with one example.

Testing/Quality Assurance Team - They will perform various tests in quality server. Test Script will be designed based on functional specification. If they find any bugs developer will revert code to development server. Once every test is passed they will give a test result document. The entire result will be given as a Excel sheet and uploaded in a tracker. The bugs are noted and resolved. The Excel sheet is known as Completion sheet or bug sheet. Once QA testing is done the code is moved to production server.

Administration - Plays a main role in many main sectors like Network, Database, Application. Network team handles issues like WiFi downtime etc. Database team handles database security and protects it from hackers. Application team handles the product in product-based companies. Security team handles security. It is the most powerful team in any office and any issues should be sorted out as soon as possible. They handle firewall protection, antivirus protection, subscriptions etc.

Product/Project Manager - They have the right to question everyone who comes under their project. They will talk to the client directly if any genuine reason is there for delay in project. They also have right to fire an employee if ther is no genuine reason for delay. They know entire scope of the project.

Sales and Marketing Team - Marketing team will induce the consumer to buy a product by advertising the product. Advertisement and specs will be handled by marketing team. Sales team will do a deal with the customer to get the product. They will give the quotation to the customer which the customer can rewrite and send to the sales team. Once payment is done the quotation is done by the sales team.

Human Resource(HR) - Bridge between employees and management. They will announce any new changes in policies to employees. They will also take employee concerns to management level. They are also responsible for hiring. They will take care of the person from hiring to resignation. Not technically but professionally.

Documentation - Every single action of the company, every single milestone, technical aspects, timeline, expected time, completed time, accurate time, financial things, sorting documents to send to government etc. will be handled by documentation.

Support Team - They will act as functional, developer, tester, sales and marketing, documentation. They will take the bending job and complete it.

Three main servers:

Development

Quality

Production

If developer completes coding they will deploy in development server

Unit testing is done in development server

After it passes tests it is sent to Quality Server

Steps in Waterfall Method:

Requirement - Gather requirements

Design - A prototype is designed

Development - Development based on designed POC

Testing - QA

Deployment - Once the project is deployed to production server the user will come into the picture

Maintenance - After deployment to production server project is officially closed. Any bugs or enhancements are handled in maintenance based on user reviews.

Day 29 – Preparation for milestone-3:

Code, document, individual ppt with individual contributions, gc library for milestone-3**.**

Day 30 – Algorithms:

A* Algorithm:

- Pathfinding algorithm which finds the shortest path from starting and goal node.
- Finds shortest path based on weightage.
- It is a heuristic search algorithm.
- It is an extension of Djikstra's algorithm.
- It works on directed graph.

Sorting Algorithms:

1. Bubble Sort
2. Selection Sort
3. Insertion Sort
4. Merge Sort
5. Quick Sort
6. Heap Sort

Day 31 – Milestone-3 Day-1:

Some members presented their progress in the projects.

Day 32 – Milestone-3 Day-2:

The remaining members presented their progress in the projects.

Day 33 – Project Preparation:

We were instructed to work on our project for milestone-4.

Day 34 – Project Work:

We were instructed to work on our project for milestone-4.

Day 35 – Reinforcement Learning:

Q-Learning:-

It is a decision making algorithm used to take the best decisions using Q values that are stored in Q tables.

Input is given and based on output and human feedback the algorithm learns and stored in agent's memory. This memory is denoted as learners. The learners will take note and based on this the model will rewrite a correct answer. 2 inputs - human input and model recorrected input.

Key components-

1. Q values and Action values
2. Rewards
3. Temporal Difference(TD)
4. Greedy Policies - Exploration and Exploitation

Q values represent the expected rewards

Rewards are given based on the action

TD will be defined based on highest Q values

Greedy has a random action. Agent will work based on random action. Greedy will be any Q value taken by the algorithm.

Q learning algorithm uses iterations to get a particular action correctly.

Q table is initialized first.

Action is what we need to execute next.

After action is performed it will measure the rewards.

Rewards are calculated based on the outcomes and conditions we are providing.

If reward value is low then Q table will not be updated.

State will auto populate several times and gives perfect answer.

An agent will be selected for an action by agent selector.

Q values are calculated using TD with Bellman's equation.

What actions need to be taken will be determined by Bellman's equation.

Day 36 – Backpropagation algorithm:

It is a basic algorithm used to adjust and recalculate weights.

Used to calculate the error of FNN.

Uses chain rule of calculus to compute the gradients.

Uses techniques like gradient descent.

It is a simple algorithm but can solve even for complex datasets.

As complexity of hidden layers increases, accuracy of detecting error increases.

Forward and backward pass

Forward pass will calculate error from input layer to hidden layer and then goes to output layer.

Backward pass does the opposite going from output layer to input layer updating weights.

Backward pass is called backpropagation.

Several optimization techniques are present.

Activation functions - Sigmoid, SoftMax, RELU etc.

Sigmoid and SoftMax are commonly used.

Activation function will introduce non-linearities.

Mean Squared Error is used to calculate the error.

What the algorithm must have:-

Efficient weight update

Scalability

Automatically adjust parameters and improve performance

Gradient Descent:

Batch Gradient Descent

Mini Gradient Descent

GED

Feed Forward Neural Networks(FNN):

We evaluate FNN using accuracy, precision, recall, f1 score and confusion matrix.

FNN can be implemented in python using TensorFlow and Keras libraries.

Day 37 – AI Algorithms:

DFS:

Depth First Search

Goes from root node to leaf node of the branch and then explores other children

BFS:

Breadth First Search

Searches all nodes in one level and then searches all nodes in next level

Depth Limited Search:

Same like DFS but has a limit to how deep it can explore

Uniform Cost Search:

Based on queue which has lowest cumulative cost it goes to that node

Travelling Salesman Problem

Day 38 – Project Preparation:

We were instructed to work on the project.

Day 39 – Project Preparation:

Milestone 4 will be on 30 and 31st dec

Complete your 100 percentage of your project

Without fail join the class on Tuesday and wednesday

Onegroup ppt

Full project code

Should be there in your group git hub

Five documents, one ppt, code in group GitHub repository

Day 40 – Milestone – 4:

In milestone 4 the completed project was presented