# Vikas Institute of Engineering and Technology, Gorakhpur

**Session 2021-22**

## Mini Project Report:
## Jarvis – Virtual Assistant

## Department of –
## Computer Science and Engineering

**Under Supervision of:**

Er. Vivek Patel (HOD)
Er. Pragya Singh
Er. Ajeet Sir
Er. Ashish Vishwakarma

**Submitted By:**

| Sr no. | Name | Task |
|---|---|---|
| 1. | Sunny Patel (Group Leader) | Project Management with Structured Programming |
| 2. | Akhilesh Yadav | Library Management |
| 3. | Sundaram Chaurasiya | Selenium, Beautiful Soup |
| 4. | Anand Kumar Yadav | Speech_Recognizer package |
| 5. | Meena Sahani | Pyttsx3, text-to-speech |
| 6. | Shivraj Sahani | Keyboad,Mouse Automation |
| 7. | Shushil Pandey | API Management |
| 8. | Udaybhan Nishad | App Testing |

# Acknowledgement

In completing this project report on project titled **JARVIS - VIRTUAL ASSISTANT**, I had to take the help and guideline of a few respected people, who deserve my greatest gratitude.

The completion of this project report gives me much Pleasure. I would like to show my gratitude **to Er. Vivek Patel** & **Er. Pragya Singh** for giving me a good guideline for project throughout numerous consultations. I would also like to expand my deepest gratitude to all those who have directly and indirectly guided us in writing this project report.

Many people, especially my classmates and friends themselves, have made valuable comments and suggestions on this proposal which gave me inspiration to improve my project. Here I thank all the people for their help directly and indirectly to complete this project report.

Sunny Patel,

# Table of Contents

# VIRTUAL ASSISTANT

## 1. INTRODUCTION

In today's era almost all tasks are digitalized. We have Smartphone in hands and it is nothing less than having world at your finger tips. These days we aren't even using fingers. We just speak of the task and it is done. There exist systems where we can say Text Dad, "I'll be late today." And the text is sent.

Virtual Assistants are software programs that help you ease your day to day tasks, such as showing weather report, news report, searching queries ,playing songs etc. They can take commands via text (online chat bots) or by voice. Voice based intelligent assistants need an invoking word or wake word to activate the listener, followed by the command. For my project the wake word is JARVIS. We have so many virtual assistants, such as Apple's Siri, Amazon's Alexa and Microsoft's Cortana. For this project, wake word was chosen JARVIS.

This system is designed to be used efficiently on desktops. Personal assistant software improves user productivity by managing routine tasks of the user and by providing information from online sources to the user. JARVIS is effortless to use. Call the wake word 'JARVIS' followed by the command. And within seconds, it gets executed.

# 1.1 BACKGROUND AND MOTIVATION

There already exist a number of desktop virtual assistants. A few examples of current virtual assistants available in market are discussed in this section along with the tasks they can provide and their drawbacks.

### SIRI from Apple

SIRI is personal assistant software that interfaces with the user thru voice interface, recognizes commands and acts on them. It learns to adapt to user's speech and thus improves voice recognition over time. It also tries to converse with the user when it does not identify the user request.

It integrates with calendar, contacts and music library applications on the device and also integrates with GPS and camera on the device. It uses location, temporal, social and task based contexts, to personalize the agent behavior specifically to the user at a given point of time.

### Supported Tasks

- Call someone from my contacts list
- Launch an application on my iPhone
- Send a text message to someone
- Set up a meeting on my calendar for 9am tomorrow
- Set an alarm for 5am tomorrow morning
- Play a specific song in my iTunes library
- Enter a new note

### Drawback

SIRI does not maintain a knowledge database of its own and its understanding comes from the information captured in domain models and data models.

### ReQall

ReQall is personal assistant software that runs on smartphones running Apple iOS or Google Android operating system. It helps user to recall notes as well as tasks within a location and time context. It records user inputs and converts them into commands, and monitors current stack of user tasks to proactively suggest actions while considering any changes in the environment. It also presents information based on the context of the user,

as well as filter information to the user based on its learned understanding of the priority of that information.

## Supported Tasks

- Reminders
- Email
- Calendar, Google Calendar
- Outlook
- Evernote
- Facebook, LinkedIn
- News Feeds

## Drawback

Will take some time to put all of the to-do items in – you could spend more time putting the entries in than actually doing the revision.

# 1.2 OBJECTIVES

Main objective of building personal assistant software (a virtual assistant) is using semantic data sources available on the web, user generated content and providing knowledge from knowledge databases. The main purpose of an intelligent virtual assistant is to answer questions that users may have. This may be done in a business environment, for example, on the business website, with a chat interface. On the mobile platform, the intelligent virtual assistant is available as a call-button operated service where a voice asks the user “What can I do for you?” and then responds to verbal input.

Virtual assistants can tremendously save you time. We spend hours in online research and then making the report in our terms of understanding. JARVIS can do that for you. Provide a topic for research and continue with your tasks while JARVIS does the research. Another difficult task is to remember test dates, birthdates or anniversaries. It comes with a surprise when you enter the class and realize it is class test today. Just tell JARVIS in advance about your tests and she reminds you well in advance so you can prepare for the test.

One of the main advantages of voice searches is their rapidity. In fact, voice is reputed to be four times faster than a written search: whereas we can write about 40 words per minute, we are capable of speaking around 150 during the same period of time15. In this respect, the ability of personal assistants to accurately recognize spoken words is a prerequisite for them to be adopted by consumers.

# 1.3 PURPOSE, SCOPE AND APPILCABILITY

## Purpose

Purpose of virtual assistant is to being capable of voice interaction, music playback, making to-do lists, setting alarms, streaming podcasts, playing audiobooks, and providing weather, traffic, sports, and other real-time information, such as news. Virtual assistants enable users to speak natural language voice commands in order to operate the device and its apps.

There is an increased overall awareness and a higher level of comfort demonstrated specifically by millennial consumers. In this ever-evolving digital world where speed, efficiency, and convenience are constantly being optimized, it's clear that we are moving towards less screen interaction.

## Scope

Voice assistants will continue to offer more individualized experiences as they get better at differentiating between voices. However, it's not just developers that need to address the complexity of developing for voice as brands also need to understand the capabilities of each device and integration and if it makes sense for their specific brand. They will also need to focus on maintaining a user experience that is consistent within the coming years as complexity becomes more of a concern. This is because the visual interface with voice assistants is missing. Users simply cannot see or touch a voice interface.

## Applicability

The mass adoption of artificial intelligence in users' everyday lives is also fueling the shift towards voice. The number of IoT devices such as smart thermostats and speakers are giving voice assistants more utility in a connected user's life. Smart speakers are the number one way we are seeing voice being used. Many industry experts even predict that nearly every application will integrate voice technology in some way in the next 5 years.

The use of virtual assistants can also enhance the system of IoT (Internet of Things). Twenty years from now, Microsoft and its competitors will be offering personal digital assistants that will offer the services of a full-time employee usually reserved for the rich and famous.

# 2. SURVEY OF TECHNOLOGY

## Python

Python is an OOPs (Object Oriented Programming) based, high level, interpreted programming language. It is a robust, highly useful language focused on rapid application development (RAD). Python helps in easy writing and execution of codes. Python can implement the same logic with as much as 1/5th code as compared to other OOPs languages.

Python provides a huge list of benefits to all. The usage of Python is such that it cannot be limited to only one activity. Its growing popularity has allowed it to enter into some of the most popular and complex processes like Artificial Intelligence (AI), Machine Learning (ML), natural language processing, data science etc. Python has a lot of libraries for every need of this project. For JARVIS, libraries used are speech-recognition to recognize voice, Pyttsx3 for text to speech, selenium for web automation etc.

Python is reasonably efficient. Efficiency is usually not a problem for small examples. If your Python code is not efficient enough, a general procedure to improve it is to find out what is taking most the time, and implement just that part more efficiently in some lowerlevel language. This will result in much less programming and more efficient code (because you will have more time to optimize) than writing everything in a low-level language.

## Selenium

Selenium is an umbrella project for a range of tools and libraries that enable and support the automation of web browsers.

It provides extensions to emulate user interaction with browser, a distribution server for scaling browser allocation, and the infrastructure for implementation of the W3C WebDriver specification that lets you write interchangeable code for all major web browsers.

This project is made possible by volunteer contributors who have put in thousand of hours of their own time, and made the source code freely available for any one to use, enjoy, and improve.

Selenium brings together browser vendors, engineers, and enthusiast to further an open discussion around automation of the web platform. The project organizes an annual conference to teach and nurture the community.

At the core of Selenium is WebDriver, an interface to write instruction sets that can be run interchangeably in many browsers.

## PyAutoGUI

PyAutoGUI lets your Python scripts control the mouse and keyboard to automate interactions with other applications. The API is designed to be simple. PyAutoGUI works on Windows, macOS, and Linux, and runs on Python 2 and 3.

PyAutoGUI has several features:

- Moving the mouse and clicking in the windows or other applications.
- Sending Keystrokes to applications(for example, to fill out forms).
- Take screenshots, and given an image(for example, of a button or checkbox), find it on the screen.
- Locate an application's window, and move, resize, maximize, minimize, or close it (Windows-only, currently).
- Display alert and message boxes.

## Pyttsx3

Pyttsx3 is a text-to-speech conversion library in Python. Unlike alternative libraries, it works offline and is compatible with both Python 2 and 3. An application invokes the pyttsx3.init() factory function to get a reference to a pyttsx3. Engine instance. it is a very easy to use tool which converts the entered text into speech.

The pyttsx3 module supports two voices first is female and the second is male which is provided by "sapi5" for windows.

It supports three TTS engines :

- sapi5 – SAPI5 on Windows
- nsss – NSSpeechSynthesizer on Mac OS X
- espeak – eSpeak on every other platform

## Speech Recognition

This is a library for performing speech recognition, with support for several engines and APIs, online and offline. It supports APIs like Google Cloud Speech API, IBM Speech to Text, Microsoft Bing Voice Recognition etc.

Speech recognition, as the name suggests, refers to automatic recognition of human speech. Speech recognition is one of the most important tasks in the domain of human computer interaction. If you have ever interacted with Alexa or have ever ordered Siri to complete a task, you have already experienced the power of speech recognition.

Speech recognition has various applications ranging from automatic transcription of speech data (like voicemails) to interacting with robots via speech.

In this tutorial, you will see how we can develop a very simple speech recognition application that is capable of recognizing speech from audio files, as well as live from a microphone. So, let's begin without further ado.

Several speech recognition libraries have been developed in Python. However we will be using the SpeechRecognition library, which is the simplest of all the libraries.

## Web Scraping

Scraping is simply a process of extracting (from various means), copying and screening of data.

When we do scraping or extracting data or feeds from the web (like from web-pages or websites), it is termed as web-scraping.

So, web scraping which is also known as web data extraction or web harvesting is the extraction of data from web. In short, web scraping provides a way to the developers to collect and analyze data from the internet.

Why Web-scraping?

- Data for research
- Product prices & popularity comparison
- SEO Monitoring
- Search Engine
- Sales Marketing

# 3. REQUIREMENT AND ANALYSIS

System Analysis is about complete understanding of existing system and finding where the existing system fails. The solution is determined to resolve issues in the proposed system. It defines the system. The system is divided into smaller parts. Their functions and inter relation of these modules are studied in system analysis. The complete analysis is followed below.

## 3.1 Problem definition

Usually, user needs to manually manage multiple sets of applications to complete one task. For example, a user trying to make a travel plan needs to check for airport codes for nearby airports and then check travel sites for tickets between combinations of airports to reach the destination. There is need of a system that can manage tasks effortlessly.

We already have multiple virtual assistants. But we hardly use it. There are number of people who have issues in voice recognition. These systems can understand English phrases but they fail to recognize in our accent. Our way of pronunciation is way distinct from theirs. Also, they are easy to use on mobile devices than desktop systems. There is need of a virtual assistant that can understand English in Indian accent and work on desktop system.

When a virtual assistant is not able to answer questions accurately, it's because it lacks the proper context or doesn't understand the intent of the question. Its ability to answer questions relevantly only happens with rigorous optimization, involving both humans and machine learning. Continuously ensuring solid quality control strategies will also help manage the risk of the virtual assistant learning undesired bad behaviors. They require large amount of information to be fed in order for it to work efficiently.

Virtual assistant should be able to model complex task dependencies and use these models to recommend optimized plans for the user. It needs to be tested for finding optimum paths when a task has multiple sub-tasks and each sub-task can have its own subtasks. In such a case there can be multiple solutions to paths, and the it should be able to consider user preferences, other active tasks, priorities in order to recommend a particular plan.

# 3.2 REQUIREMENT SPECIFICATION

Personal assistant software is required to act as an interface into the digital world by understanding user requests or commands and then translating into actions or recommendations based on agent's understanding of the world.

JARVIS focuses on relieving the user of entering text input and using voice as primary means of user input. Agent then applies voice recognition algorithms to this input and records the input. It then use this input to call one of the personal information management applications such as task list or calendar to record a new entry or to search about it on search engines like Google, Bing or Yahoo etc. Focus is on capturing the user input through voice, recognizing the input and then executing the tasks if the agent understands the task. Software takes this input in natural language, and so makes it easier for the user to input what he or she desires to be done.

Voice recognition software enables hands free use of the applications, lets users to query or command the agent through voice interface. This helps users to have access to the agent while performing other tasks and thus enhances value of the system itself. JARVIS also have ubiquitous connectivity through Wi-Fi or LAN connection, enabling distributed applications that can leverage other APIs exposed on the web without a need to store them locally.

Virtual assistants must provide a wide variety of services. These include:

- Providing information such as weather, facts from e.g. Wikipedia etc.
- Set an alarm or make to-do lists and shopping lists.
- Remind you of birthdays and meetings.
- Play music from streaming services such as Saavn and Gaana.
- Play videos, TV shows or movies on televisions, streaming from e.g. Netflix or Hotstar.
- Book tickets for shows, travel and movies.

## Feasibility Study

Feasibility study can help you determine whether or not you should proceed with your project. It is essential to evaluate cost and benefit. It is essential to evaluate cost and benefit of the proposed system. Three types of feasibility study are taken into consideration.

1. **Technical feasibility:** It includes finding out technologies for the project, both hardware and software. For virtual assistant, user must have microphone to convey their message and a speaker to listen when system speaks. These are very cheap now a days and everyone generally possess them. Besides, system needs internet connection. While using JARVIS, make sure you have a steady internet connection. It is also not an issue in this era where almost every home or office has Wi-Fi.

2. **Economical feasibility:** Here, we find the total cost and benefit of the proposed system over current system. For this project, the main cost is documentation cost. User also would have to pay for microphone and speakers. Again, they are cheap and available. As far as maintenance is concerned, JARVIS won't cost too much.

3. **Organizational feasibility:** This shows the management and organizational structure of the project. This project is built by a team. The management tasks are all to be carried out by a few team members. That won't create any management issues and will increase the feasibility of the project.

# 3.3 HARDWARE AND SOFTWARE REQUIREMENTS

The software is designed to be light-weighted so that it doesn't be a burden on the machine running it. This system is being build keeping in mind the generally available hardware and software compatibility. Here are the minimum hardware and software requirement for virtual assistant.

## Hardware:

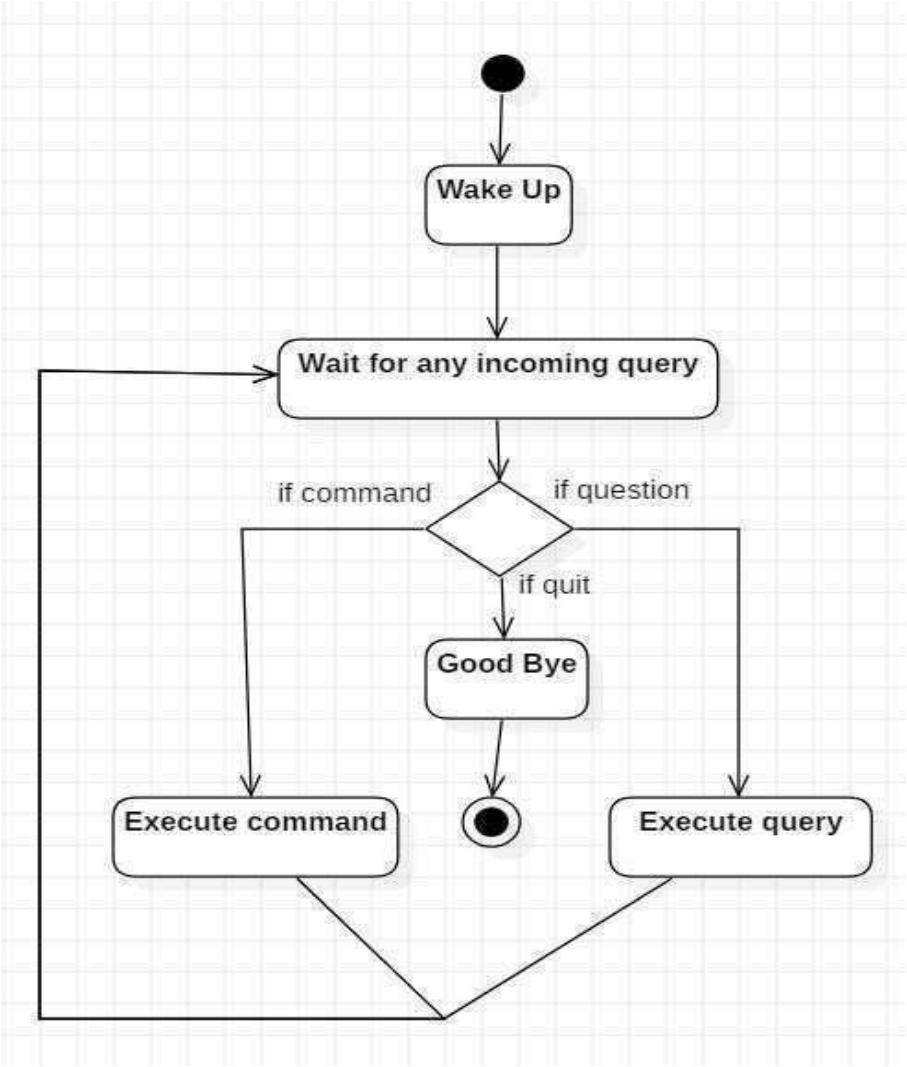
- Pentium-pro processor or later.
- RAM 512MB or more.

## Software:

- Windows 7(32-bit) or above.
- Python 2.7 or later
- Chrome Driver
- Selenium Web Automation

# 4. SYSTEM DESIGN

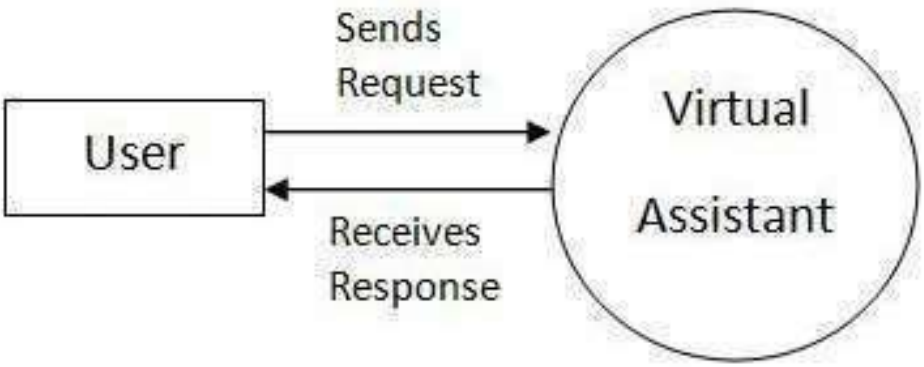
## 4.1 ACTIVITY DIAGRAM



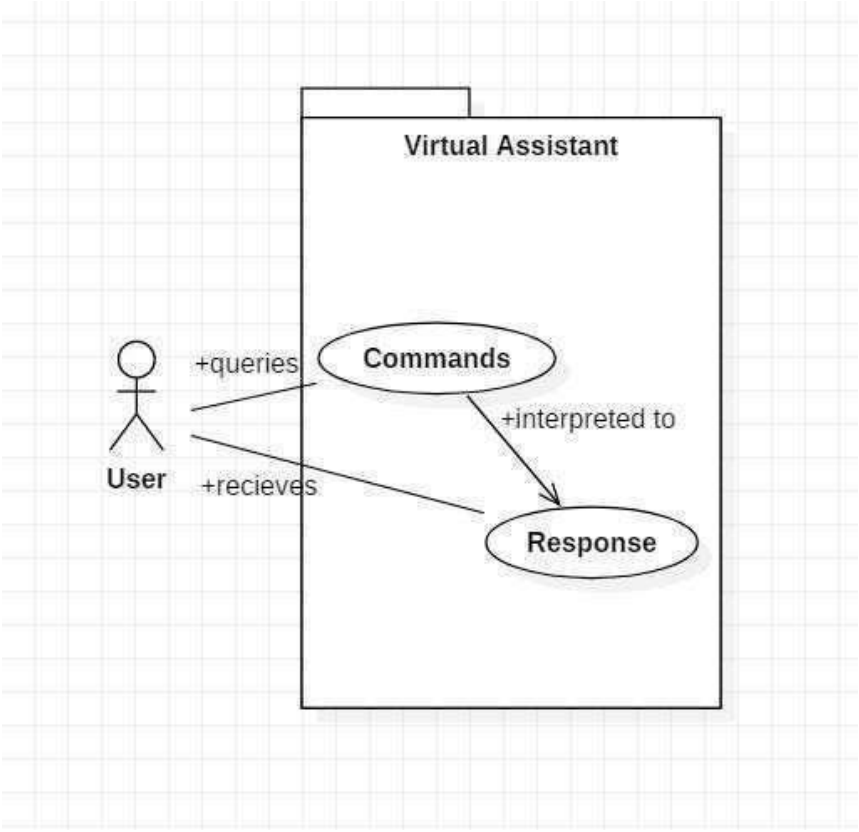

Initially, the system is in idle mode. As it receives any wake up call it begins execution. The received command is identified whether it is a questionnaire or a task to be performed. Specific action is taken accordingly.

After the Question is being answered or the task is being performed, the system waits for another command. This loop continues unless it receives quit command. At that moment, it goes back to sleep.

# 4.2 DATA FLOW DIAGRAM

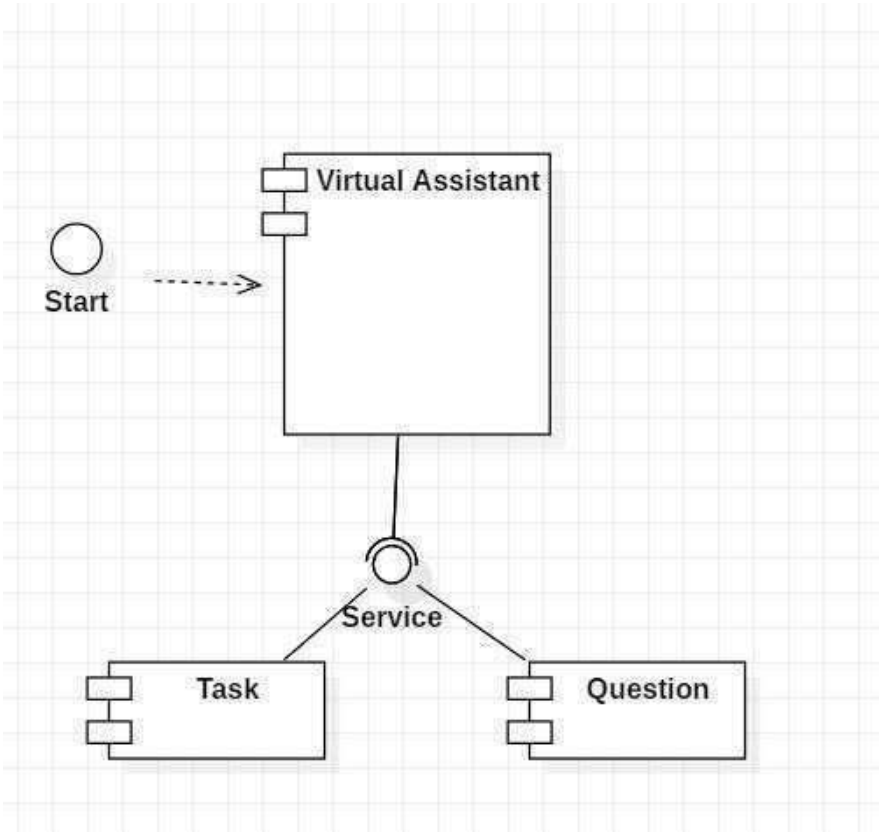



# 4.3 USE CASE DIAGRAM




As you can see in this project there is only one user. The user queries command to the system. System then interprets it and fetches answer. The response is sent back to the user.

# 4.4 COMPONENT DIAGRAM




The main component here is the Virtual Assistant. It provides two specific service, executing Task or Answering your question.

# 5. Integrated Development Environment(IDE)

An integrated development environment (IDE) is software for building applications that combines common developer tools into a single graphical user interface (GUI).

## Visual Studio Code (IDE)

We have used Visual Studio Code for our project development, it is an integrated development environment (IDE) from Microsoft. It is used to develop computer programs, as well as websites, web apps, web services and mobile apps. Although the editor is relatively lightweight, it includes some powerful features that have made VS Code one of the most popular development environment tools in recent times.

It is a lightweight but powerful source code editor which runs on your desktop and is available for Windows, macOS and Linux. It comes with built-in support for JavaScript, TypeScript and Node.js and has a rich ecosystem of extensions for other languages (such as C++, C#, Java, Python, PHP, Go) and runtimes (such as .NET and Unity).

At its heart, Visual Studio Code features a lightning fast source code editor, perfect for day-to-day use. With support for hundreds of languages, VS Code helps you be instantly productive with syntax highlighting, bracket-matching, auto-indentation, box-selection, snippets, and more. Intuitive keyboard shortcuts, easy customization and communitycontributed keyboard shortcut mappings let you navigate your code with ease.

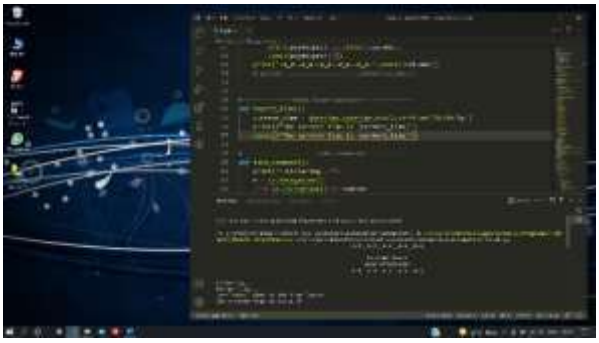
For serious coding, you'll often benefit from tools with more code understanding than just blocks of text. Visual Studio Code includes built-in support for IntelliSense code completion, rich semantic code understanding and navigation, and code refactoring.

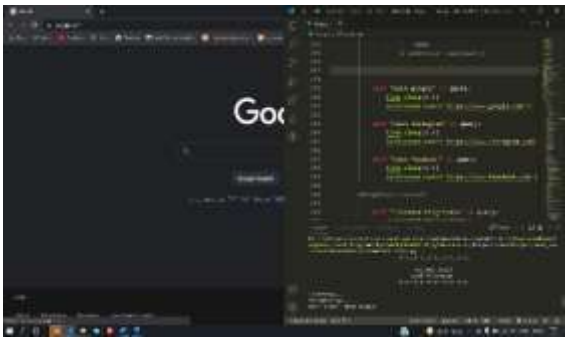
And when the coding gets tough, the tough get debugging. Debugging is often the one feature that developers miss most in a leaner coding experience, so we made it happen. Visual Studio Code includes an interactive debugger, so you can step through source code, inspect variables, view call stacks, and execute commands in the console.

VS Code also integrates with build and scripting tools to perform common tasks making everyday workflows faster. VS Code has support for Git so you can work with source control without leaving the editor including viewing pending changes diffs.

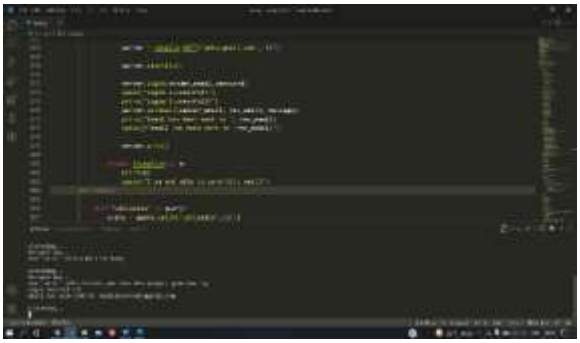
# 6. PROJECT SCREENSHOTS

**Screenshot :** Telling time & Search Google

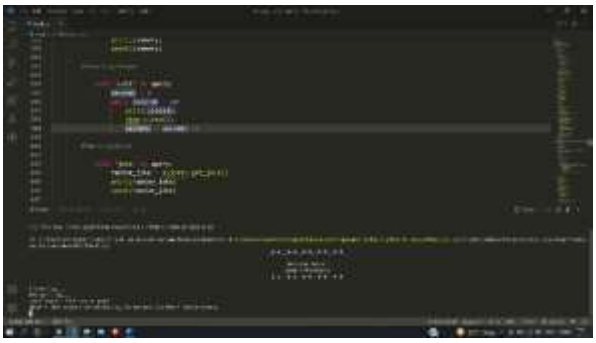




**Screenshot:** Opening Instagram & Sending Email



**Screenshot :** Wikipedia Search & Telling Joke



**Screenshot :** Playing Music


and many more…

# REFERENCE AND BIBLIOGRAPHY


- Websites referred
  - https://www.google.com/
  - https://stackoverflow.com/
  - https://www.w3schools.com/
  - https://www.python.org/

- Books referred
  - Python Programming – Adam Stewart.
  - Learning Python - Mark Lutz.

- YouTube Channels referred
  - CodeWithHarry
  - freeCodeCamp

- Documents referred
  - Create Virtual Assistant - Teams | Microsoft Docs
  - https://en.wikipedia.org/wiki/Virtual_assistant /


THANK YOU