**REPORT ON**

Career Recommendation System

*Submitted in partial fulfilment of the requirements for the award of the degree of*

**BACHELOR OF COMPUTER APPLICATIONS**

****

**Batch: 2021 - 24**

***Under the Guidance of Submitted By***

Dr. Ruchi Sawhney Ankit Alex Minz 04328402021

Dibya Bharat 00928402021

**Bosco Technical Training Society**

**Don Bosco Technical School, Okhla Road, New Delhi-110025**

**9643868820, 8527787221, 011-41033889**

**Affiliated to**

****

**Guru Gobind Singh Indraprastha University**

### Sector - 16C, Dwarka, New Delhi - 110078 (India)

### Phone: +91-11-25302170, Fax: +91-11- 25302111

### E-Mail Id: ggsipu.pr@rediffmail.com, pro@ipu.ac.in

# 

## 

## Acknowledgement

The note starts with thanks to Almighty who actually created this piece of work and helped us when things were not easy for us.

I am very grateful and indebted to my Faculty/Guide <<Guide Name>> who immensely helped and rendered her/his valuable advice, precious time, knowledge and relevant information regarding the collection of material. She/He has been a major source of inspiration throughout the project as he not only guided me throughout this Project Report <<Project Title>> but also encouraged me to solve problems that arose during this report.

His guidance and suggestions about this Project report have really enlightened me. It has been a great help to support to have him around.

And finally, I would like to mention appreciation to our parents and friends who have been instrumental throughout this period by providing unrelenting encouragement.

Name of Students       Enrolment No. Signature

**CERTIFICATE**

This is to certify that the dissertation/project report entitled “Career Recommendation System” done by me is an authentic work carried out for the partial fulfilment of the requirements for the award of the degree of Bachelor of Computer Applications under the guidance of Mrs Ruchi Sawhney. The matter embodied in this project work has not been submitted earlier for award of any degree or diploma to the best of my knowledge and belief.

Signature of the student

ANKIT ALEX MINZ

04328402021

MAIN REPORT

Objective and Scope

**Objectives:**

* Develop a machine learning-driven recommendation system that analyzes students' academic performance, interests, strengths, and career aspirations which provides course, stream, and career recommendations based on content filtering.
* Identify and address subject-specific weaknesses that may hinder students' chosen career paths.
* Implement a user-friendly questionnaire to refine recommendations based on user tendencies and make further recommendations on things of their interest that they lack awareness of or even provide new things to get interested in for those who have exhausted their option.
* Suggest improvement strategies: For subjects in which students are weak, the system will offer tailored recommendations on strengthening their skills and improving their performance.

**Scope:**

* The project's scope encompasses providing students with informed career recommendations following their 10th-grade examinations.
* Recommendations are personalized through an in-depth analysis of academic data, interests, and subject weaknesses.
* Strategies to enhance proficiency in relevant subjects are offered.
* A user-centric questionnaire refines recommendations by evaluating user inclinations.

Theoretical Background Definition of Problem

Theoretical Background:

**Python:**

Python is a high-level, general-purpose, and versatile programming language. Python programming language is being used in web development, Machine Learning applications, along with all cutting-edge technology in Software Industry. Python’s Background in Desktop Application Development and Machine Learning comes from its cross platform nature and various libraries for Example: Pyqt5, Pandas, Numpy, Sci-kit Learn etc.

**Pyqt5:**

There are so many options provided by Python to develop GUI application and PyQt5 is one of them. PyQt5 is cross-platform GUI toolkit, it has a set of python bindings for Qt v5. One can develop an interactive desktop application with much ease because of the tools and simplicity provided by this library. A GUI application consists of Front-end and Back-end. PyQt5 has provided a tool called ‘QtDesigner’ to design the front-end by drag and drop method so that development can become faster and one can give more time on back-end stuff.

Due to the nature of Desktop Application development and Pyqt5 , Event Driven Programming come into play. It is a programming paradigm where different parts of the application or entities communicate by sending messages to one another through an intermediary. In the context of Gui development and desktop application development this refers to different components reacting to each other through signals and events written out by the programmer. For example, a button being clicked in a Pyqt5 gui interface is connected to function which would then be executed.

**Scikit-Learn:**

Scikit-learn is a powerful open-source library in Python used for machine learning tasks such as classification, regression, clustering, and more. It is built on top of other popular scientific computing libraries, such as NumPy, SciPy, and Matplotlib. It offers a wide range of tools for machine learning and statistical modelling, including various algorithms and utilities for data preprocessing, model selection, evaluation, and data visualization.

In practical terms, scikit-learn serves as a powerful toolbox for tasks such as data preprocessing, model training, and model evaluation. It includes utilities for handling data, implementing feature selection, and assessing model performance through cross-validation and grid search.

**Pandas:**

Pandas is an open-source Python library used for data manipulation and analysis. It provides powerful data structures and tools for working with structured data, primarily in the form of DataFrame objects. Pandas is widely utilized in machine learning for its ability to handle various data formats, such as CSV files, Excel sheets, SQL databases, and more. Its main data structure, the DataFrame, allows for easy indexing, slicing, cleaning, transforming, and aggregating data, making it an essential tool for data preprocessing in machine learning workflows.

It enables practitioners to load datasets, handle missing values, perform feature engineering by creating new features or transforming existing ones, encode categorical variables, and split data into training and testing sets. The ability to manipulate and prepare data efficiently using pandas is fundamental for ensuring that machine learning models receive high-quality input, leading to more accurate predictions or classifications.

**Numpy:**

NumPy, short for Numerical Python, is a fundamental open-source library in Python for numerical computing. It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays efficiently. In machine learning, NumPy is extensively used for data manipulation and preparation. It plays a central role in handling datasets, transforming features, and performing mathematical operations required during the preprocessing phase. Many machine learning libraries, including scikit-learn, TensorFlow, and PyTorch, leverage NumPy arrays as the underlying data structure. NumPy's efficient implementation of array operations significantly speeds up computations, making it an essential tool for implementing algorithms and models efficiently in machine learning workflows.

**Machine Learning:**

Machine Learning is a branch of artificial intelligence that develops algorithms by learning the hidden patterns of the datasets used it to make predictions on new similar type data, without being explicitly programmed for each task.   
Machine learning is used in many different applications, from image and speech recognition to natural language processing, recommendation systems, fraud detection, portfolio optimization, automated task, and so on. Machine learning models are also used to power autonomous vehicles, drones, and robots, making them more intelligent and adaptable to changing environments.  
A typical machine learning tasks are to provide a recommendation. Recommender systems are a common application of machine learning, and they use historical data to provide personalized recommendations to users.  
Types of Machine Learning:

1. Supervised Machine Learning
2. Unsupervised Machine Learning
3. Reinforcement Machine Learning

**Supervised Learning:**

In this project supervised learning has been chosen for usage, supervised learning is a type of machine learning in which the algorithm is trained on the labeled dataset i.e., the data may have correct values as well. It learns to map input features to targets based on labeled training data. In supervised learning, the algorithm is provided with input features and corresponding output labels, and it learns to generalize from this data to make predictions on new, unseen data.

There are two main types of supervised learning:

1. Regression
2. Classification

**Classification:**

Classification is a type of supervised learning where the algorithm learns to assign input data to a specific category or class based on input features. The output labels in classification are discrete values. Classification algorithms can be binary, where the output is one of two possible classes, or multiclass, where the output can be one of several classes. The different Classification algorithms in machine learning are: Logistic Regression, Naive Bayes, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), etc.

**KNN Algorithm:**

In this project the KNN algorithm has been chosen from classification type supervised learning, K-Nearest Neighbours is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining, and intrusion detection. It can also handle both numerical and categorical data, making it a flexible choice for various types of datasets in classification and regression tasks. It is a non-parametric method that makes predictions based on the similarity of data points in a given dataset. K-NN is less sensitive to outliers compared to other algorithms. The applications of KNN are data preprocessing, pattern recognition, Recommendation Engines.

**Content based filtering:**

Content-based filtering is a machine learning approach used in recommendation systems to suggest items to users based on the characteristics of the items and the preferences of the users. Instead of relying on the collaborative behaviour of users (as seen in collaborative filtering), content-based filtering considers the features or attributes of items and recommends new items that are similar to those a user has liked or interacted with in the past.

In machine learning applications, content-based filtering finds its use in various recommendation systems, such as suggesting movies, music, articles, or products to users based on their historical preferences or explicit feedback.

Using content based filtering in machine learning models can recommend new items that align with users' tastes, making the recommendation process more personalized and potentially increasing user satisfaction and engagement.

**Problem Statement:**

**Problem:**

The "Career Recommendation System" project addresses a pressing issue in the educational landscape, primarily concerning students transitioning from their 10th-grade examinations. At this critical juncture, students often find themselves at a crossroads, faced with the formidable task of choosing academic courses or streams that will significantly shape their future careers. However, the educational system currently lacks a structured and comprehensive decision-making framework to guide students through this crucial phase of their academic journey.

In particular, students grapple with several challenges:

* **Lack of Systematic Guidance:** The absence of a systematic guidance system leaves many students in the dark, unsure of which path to follow. Without clear direction, they may make ill-informed decisions that can impact their career prospects for years to come.
* **Balancing Performance and Aspirations:** Striking the right balance between their academic performance, personal interests, and inherent strengths is a complex and often bewildering task. Students must align their educational choices with their career aspirations, a process that is far from straightforward.

As a result, many students are left navigating this critical decision-making process with minimal support or personalized advice. This lack of guidance can have profound consequences on their educational and career trajectories, potentially leading to mismatches between their chosen paths and their true potential.

**Background:**

The existing educational framework has long grappled with the challenge of equipping students with the tools and knowledge necessary to make informed career choices. Traditionally, students have relied on a mix of limited resources, including career counsellors and educational advisors, which often fall short of providing the personalized guidance needed.

The complexities of aligning academic performance, personal interests, and career goals further exacerbate the problem. The evolving job market demands that students choose their paths with care, as they prepare for careers in fields that are continually changing and diversifying. Consequently, there is a growing need for an innovative solution that addresses these challenges effectively.

**Relevance:**

The "Career Recommendation System" project is profoundly relevant in the context of today's educational landscape. It directly addresses the critical issue of career decision-making for students, offering a tailored and data-driven approach to support their academic and vocational aspirations. Several key aspects underline the system's relevance:

* **Personalized Recommendations:** By leveraging data on academic performance, personal interests, and strengths, the system provides personalized course and career recommendations. This tailoring ensures that students are directed toward paths that genuinely resonate with their capabilities and aspirations.
* **Subject Weakness Mitigation:** The system goes a step further by diagnosing subject-specific weaknesses. This critical feature empowers students to address these weaknesses and bolster their educational foundations, thereby enhancing their chances of success in their chosen careers.
* **Improved Academic and Career Trajectories:** The project holds the potential to significantly improve students' academic and career trajectories. By offering precise, data-driven recommendations, it helps students make more informed decisions, reducing the likelihood of misaligned choices.

System Analysis & Design

**Use Case:**

A use case diagram is used to represent the dynamic behaviour of a system. It encapsulates the system's functionality by incorporating use cases, actors, and their relationships. It models the tasks, services, and functions required by a system/subsystem of an application. It depicts the high-level functionality of a system and also tells how the user handles a system.

It invokes persons, use cases, and several things that invoke the actors and elements accountable for the implementation of use case diagrams.

Following are the purposes of a use case diagram given below:

1. It gathers the system's needs.

2. It depicts the external view of the system.

3. It recognizes the internal as well as external factors that influence the system.

4. It represents the interaction between the actors.

**Components of use case diagram**

**Actors:** Actors represent the users or external entities interacting with the system. An actor can be a human user, another system, or any entity that interacts with the system. Actors are depicted as stick figures or named entities outside the system boundary.

**Use Cases:** Use cases represent the specific functionalities or services that the system provides to its users. Each use case describes a set of actions or interactions between the user and the system to achieve a specific goal. Use cases are depicted as ovals or ellipses within the system boundary.

**Relationships and Associations:**

**Association (Communication Line):** Lines connecting actors to use cases represent interactions between actors and use cases. These lines show that an actor is involved in or can trigger a particular use case.

**Inclusion (Include):** Indicates that one use case includes the functionality of another use case. For example, a "Login" use case might include an "Authenticate" use case.

**Extension (Extend):** Shows that one use case extends or adds optional behaviour to another use case under certain conditions. It signifies optional and conditional behaviour.

**Generalization:** Represents a relationship between more general and more specific use cases, showing inheritance or specialization. It's similar to inheritance in object-oriented programming.

**System Boundary:** The boundary or box around the use cases represents the scope or boundary of the system being modeled. It encapsulates all the use cases and actors involved within the system.



**Iterative model**

The iterative model in software engineering is an approach where the project is broken down into smaller cycles or iterations. Each iteration goes through the entire software development life cycle (SDLC), encompassing planning, design, implementation, testing, and deployment. Here's a breakdown of the steps in the iterative model:

**1. Planning:** Define the project scope, objectives, and goals. Identify the features or functionalities to be developed in the current iteration. Plan resources, timelines, and iteration goals.

**2. Requirement Analysis:** Gather and analyze user requirements specific to the current iteration. Focus on understanding what users need for the features or functionalities planned for this cycle.

**3. Design:** Create a design based on the gathered requirements. This includes architectural design, system design, and detailed design of individual components or features targeted for the current iteration.

**4. Implementation/Coding:** Develop the software based on the design specifications. Write code for the identified features or functionalities planned for this iteration.

**5. Testing:** Perform testing activities such as unit testing, integration testing, and system testing. Ensure that the developed features meet the requirements and function correctly.

**6. Evaluation/Review**: Evaluate the results of the iteration. Collect feedback from stakeholders, users, and testing. Identify areas for improvement and assess whether the iteration met its goals.

**7. Deployment:** If the iteration meets the quality standards and requirements, deploy the developed features. This might involve releasing the software to a limited user base or integrating it into the existing system.

**8. Feedback/Refinement:** Gather feedback from users and stakeholders about the deployed features. Use this feedback to refine and improve the product. This input often feeds back into the planning phase for subsequent iterations.

**9. Iterations:** Repeat the cycle for the next iteration. Incorporate changes based on feedback and update the plan to address new requirements or modifications.

**Activity Diagram**

A UML Activity diagram is a graphical representation used to model workflows and business processes within a system. It primarily focuses on the flow of activities or actions performed within a system, showcasing the sequence, conditions, and parallelism of these actions. Here are the main components of a UML Activity diagram:

1. Initial Node: Denotes the starting point of the activity diagram. It represents the initiation of the workflow.

2. Activity or Action: Represents a specific task or action within the system. It can be any operation, such as a calculation, decision-making process, or interaction.

3. Control Flow: Indicates the flow or sequence of activities. It's represented by arrows connecting different activities, showing the order in which actions occur.

4. Decision or Merge Node: Represents a branching point in the workflow where a decision is made or where different flows merge back together.

5. Fork and Join Nodes: Fork nodes split the flow into multiple concurrent flows, while join nodes synchronize these concurrent flows back into a single flow.

6. Final Node: Denotes the end of the activity diagram, representing the completion of the workflow.

7. Guard Condition: Represents the condition that determines which path to follow in a decision point. It's often depicted as text near the decision node or control flow arrow.

8.Swimlanes (Partition): Used to group activities performed by different actors or system components. They visually separate actions performed by different entities within the system.

9. Object Nodes: Represent objects or data consumed or produced during activities.

##Testing##

Software Testing is a method to assess the functionality of the software program. The process checks whether the actual software matches the expected requirements and ensures the software is bug-free. The purpose of software testing is to identify the errors, faults, or missing requirements in contrast to actual requirements. It mainly aims at measuring the specification, functionality, and performance of a software program or application.

Types of software testing:

1. Manual Testing

2. Automation Testing

##Manual Testing##

Manual testing includes testing software manually, i.e., without using any automation tool or script. In this type, the tester takes over the role of an end-user and tests the software to identify any unexpected behaviour or bug. There are different stages for manual testing such as unit testing, integration testing, system testing, and user acceptance testing. Testers use test plans, test cases, or test scenarios to test software to ensure the completeness of testing. Manual testing also includes exploratory testing, as testers explore the software to identify errors in it.

Types of manual testing:

1. White Box

2. Black Box

3. Grey Box

##White Box##

White box testing techniques analyse the internal structures the used data structures, internal design, code structure, and the working of the software rather than just the functionality as in black box testing. It is also called glass box testing or clear box testing or structural testing. White Box Testing is also known as transparent testing or open box testing.

White box testing is a software testing technique that involves testing the internal structure and workings of a software application. The tester has access to the source code and uses this knowledge to design test cases that can verify the correctness of the software at the code level.

White box testing is also known as structural testing or code-based testing, and it is used to test the software’s internal logic, flow, and structure. The tester creates test cases to examine the code paths and logic flows to ensure they meet the specified requirements.

##Input & Output Screen Design##

#---PyQt5--#

PyQt5 is cross-platform GUI toolkit, a set of python bindings for Qt v5. One can develop an interactive desktop application with so much ease because of the tools and simplicity provided by this library. A GUI application consists of Front-end and Back-end. PyQt5 has provided a tool called ‘QtDesigner’ to design the front-end by drag and drop method so that development can become faster and one can give more time on back-end stuff.

Here are some key aspects of PyQt5:

Cross-platform: Applications developed with PyQt5 can run on various operating systems like Windows, macOS, Linux, etc., without many modifications. This cross-platform capability is due to Qt's nature.

GUI Development: PyQt5 provides tools and classes to design and build graphical user interfaces using Qt Designer or programmatically within Python code. It includes widgets, layouts, dialogs, and more for creating interactive interfaces.

Event Handling: It allows developers to handle user interactions, events, signals, and slots effectively. This enables the creation of responsive applications by linking user actions to specific functions or methods.

Integration with Python: PyQt5 seamlessly integrates with Python, allowing developers to leverage the ease of Python programming while harnessing the power of the Qt framework.

Support for Multimedia: PyQt5 supports multimedia functionalities, such as playing audio and video files, thanks to Qt's multimedia modules.

Database Integration: It provides tools for integrating with databases, allowing developers to create applications that interact with various database systems

#--Components of pyqt5--#

PyQt5 offers a range of components and classes that enable developers to create Graphical User Interfaces (GUIs) in Python. Here are some key components commonly used in PyQt5 for GUI creation:

Widgets: PyQt5 provides a variety of widgets that form the building blocks of a GUI. These include:

QPushButton: Button widget for triggering actions.

QLabel: Display text or an image.

QLineEdit: Single-line text input field.

QTextEdit: Multiline text input/editing field.

QCheckBox and QRadioButton: Checkboxes and radio buttons for user selections.

QComboBox and QListWidget: Dropdown lists and list widgets for item selection.

QSlider and QSpinBox: Widgets for selecting numerical values.

Layout Management: PyQt5 includes layout classes that help in organizing and arranging widgets within a window or container. Common layout classes are:

QVBoxLayout and QHBoxLayout: Vertical and horizontal box layouts.

QGridLayout: Grid-based layout for arranging widgets in rows and columns.

QFormLayout: Organizes input fields and their labels in a structured form.

Dialogs: PyQt5 provides pre-built dialog windows for specific purposes, such as file dialogs, message boxes, input dialogs, and more. These include:

QFileDialog: Dialog for file selection.

QMessageBox: Dialog for displaying messages or alerts.

QInputDialog: Dialog for user input.

Custom Widgets: Developers can create custom widgets by subclassing existing PyQt5 widgets or by creating entirely new ones to suit specific application needs.

Events and Signals: PyQt5 uses a signals and slots mechanism to handle events and inter-widget communication. This allows actions in one widget to trigger specific functions or methods in another widget.

Graphics and Multimedia: PyQt5 includes classes for working with graphics and multimedia elements, allowing developers to incorporate images, videos, and graphics within their applications.

#--Qt Designer--#

Qt Designer is a visual design tool provided by the Qt framework for creating graphical user interfaces (GUIs). It is part of the Qt development environment and is commonly used in conjunction with PyQt, a set of Python bindings for the Qt framework. Here are the key features of Qt Designer:

1. Visual Design: Qt Designer allows developers to design GUIs visually by dragging and dropping UI components onto a form. This provides a WYSIWYG (What You See Is What You Get) environment, allowing designers and developers to see the appearance of the UI as they design it.

2. Widget Box: Qt Designer provides a widget box that contains a variety of standard UI components (widgets) such as buttons, labels, text fields, and more. Developers can easily select and place these widgets on the form.

3. Layout Management: It supports the design of complex layouts through various layout managers, including vertical and horizontal box layouts, grid layouts, and form layouts. This helps in arranging and organizing widgets within a window.

4. Property Editor: Developers can use the property editor to customize the properties of UI components, such as their size, position, text, and appearance. This allows for fine-tuning the visual aspects of the GUI.

5. Signal and Slot Editing: Qt Designer provides an interface for connecting signals and slots, which facilitates event handling and interaction between different UI components. This is a fundamental aspect of Qt's programming model.

6. Integration with Code Editors: While Qt Designer is primarily a visual tool, it seamlessly integrates with code editors. Developers can easily switch between the visual design in Qt Designer and the corresponding code in their preferred integrated development environment (IDE).

7. Custom Widget Integration: Qt Designer allows developers to integrate custom widgets into the design process. This is particularly useful when developers have created their own specialized UI components.

8. Preview Mode: Developers can preview the appearance and behaviour of the GUI directly within Qt Designer before generating the code. This helps in identifying any design issues before deploying the application.

9. UI File Generation: Qt Designer saves the designed UI in a .ui file format, which is an XML-based file. This file can be loaded and used by the application code, enabling the separation of UI design and logic.