



INTERNATIONAL RESEARCH JOURNAL OF HUMANITIES AND INTERDISCIPLINARY STUDIES

(Peer-reviewed, Refereed, Indexed & Open Access Journal)

DOI : 03.2021-11278686

ISSN : 2582-8568

IMPACT FACTOR : 7.560 (SJIF 2024)

Analyzing Students' Interest in the Information Technology Sector Using Data Science Techniques

Prof. Sucheta Nitin Murumkar

Smt. Mathubai Garware Kanya Mahavidyalaya,
Sangli (Maharashtra, India)

E-mail: suchetamuurmkar7@gmail.com

DOI No. **03.2021-11278686**

DOI Link :: <https://doi-ds.org/doi/10.2024-53646381/IRJHISIC2401011>

ABSTRACT:

The information technology (IT) sector is a rapidly evolving field that plays a crucial role in shaping our digital future. Understanding students' interest in the IT sector is essential for educational institutions, policymakers, and industry professionals. This research paper aims to analyze students' interest in the information technology sector using data science techniques. By applying data science methodologies to collected data, I can gain valuable insights into the factors that influence students' interest in the IT sector, identify trends, and provide recommendations for fostering student engagement and promoting career opportunities in this field. To find out more about students' interest in it sector a well-designed questionnaire in Google Forms is used to poll students from XI, XII and Graduate Students.

Keywords: IT Sector, Data Science, Data Science Techniques, Career Opportunities, Students Interes, Python

I. Introduction:

The information technology (IT) industry is a pillar of innovation and growth in the quick-moving digital era, propelling innovations that influence every facet of contemporary life. IT offers a wide range of opportunities and challenges that pique the curiosity and ambitions of young brains, from artificial intelligence and big data to cyber security and cloud computing. In order to develop a trained workforce for the future and support technological advancement, it is essential to understand the elements that affect students' interest in the information technology sector.

With the help of data science tools, this study paper aims to delve into the complex world of students' interest in the information technology industry. We seek to understand the underlying trends, through the prism of data analysis.

A source predicts that by the end of 2023, India would have the greatest pool of software engineers. The number of software engineers hired by numerous MNCs, including Infosys, TCS,

Wipro, and others, has grown over time.

All-around software engineering businesses are reaching out to millennials with profitable career paths that include the possibility of remote working, job flexibility, career progression, rewards, and bonuses. This is due to the growing trend of tech-friendliness in the workforce. Those who wish to work in this field are drawn to it.

Additionally, as a result of the quick improvements in technology, there is a continued shortage of qualified IT personnel. Understanding the variables that draw students to the IT profession becomes increasingly important as more industries adopt digital transformation in order to fill skill gaps and maintain a consistent flow of capable and motivated workers to advance the sector.

The use of data science techniques offers a fresh strategy for doing this investigation since it allows us to precisely evaluate massive volumes of data and find relationships that would not be seen using more conventional methodologies. We seek to extract meaningful insights from complicated datasets by leveraging the strengths of machine learning techniques and statistical modelling. Data visualizations and statistical assistance will be used to show the findings in a clear and thorough manner.

2. Research Objectives:

- To assess students' levels of interest in pursuing a career or further study in the IT sector.
- To identify the factors that influence students' interest in the different domains within the IT sector, such as software development, data analytics, cybersecurity, artificial intelligence, etc.
- To explore the relationship between students' demographic characteristics (e.g., gender, educational background) and their interest in the IT sector.
- To investigate the awareness and knowledge of students regarding the IT sector, its applications, and potential career opportunities.
- To analyze the impact of educational experiences, such as coursework, internships, or extracurricular activities, on students' interest in the IT sector.
- To utilize data science techniques, including descriptive statistics, correlation analysis, and potentially predictive modelling, to gain insights into students' interest in the IT sector.
- To provide practical recommendations for educational institutions, policymakers, and industry professionals to enhance students' interest and engagement in the IT sector

III. Data science:

Data science is an interdisciplinary field that combines techniques from statistics, computer science, and domain expertise to extract valuable insights and knowledge from data. It plays a crucial role in modern research by providing powerful tools and methodologies to analyze large and complex datasets. Some common data science techniques-

- Collecting Data and Preparation
- Exploratory Data Analysis (EDA)
- Statistical Analysis:
- Machine Learning and Predictive Modeling.
- Natural Language Processing (NLP):
- Big Data and Cloud Computing:
- Ethical Considerations
- Reproducibility
- Interdisciplinary Collaboration
- Innovation and Advancements:
- These statistical techniques are just a few examples of the diverse tools data scientists employ to extract valuable insights and knowledge from data. The choice of technique depends on the type of data, research question, and the goals of the analysis.

IV. Python:

In this paper, we are going to introduce the characteristics of Python. Python is a general-purpose, high-level programming language which is widely used in the recent times ^{[1][2][3]}. Its design philosophy emphasizes code readability, and its syntax allows programmers to express concepts in fewer lines of code than would be possible in languages such as C ^[4]

Statistical analysis of data refers to the extraction of some useful knowledge from vague or complex data. Python is widely used for statistical data analysis by using data frame objects such as pandas. Statistical analysis of data includes importing, cleaning, transformation, etc. of data in preparation for analysis. The dataset of the CSV file is considered to be analyzed by python libraries which process every data from preprocessing to end result. Some libraries in python are effectively used like pandas, statsmodels, seaborn, etc that use to handle the analysis of such data. Python does data representation, data comparison, data visualization, data plotting, data testing, indexing, alignment, handling missing data, etc. Such operations are useful in data analyses that are handled by various libraries of python. Python utilizes the analysis of complex data with mix statistics with image analysis or text mining.

Some features of Python are as listed below:

- Open source and free
- Interpreted language
- Dynamic typesetting
- Portable
- Numerous IDE

Spyder:

For scientists, engineers, and data analysts, Spyder is a Python scientific environment. It combines scientific-like data exploration, interactive execution, deep inspection, and visualization skills with the extensive editing, analysis, debugging, and profiling features of a complete development tool. Spyder is funded by private donors as well as open-source supporters QuanSight and NumFOCUS. This Python IDE is small, open source, and free. It is only intended for data scientists and analysts and is entirely written in Python. When compared to other IDEs, its interface is relatively simple, but it offers every element we need in a coding environment. It includes an IPython console, file explorer, variable explorer, and text editor all in one window.

V. Scope and Methodology of the Study:

A. Scope:

The scope of this research paper is to analyze students' interest in the Information Technology (IT) sector using data science techniques. The study focuses on understanding the factors that influence students' interest in various areas of IT, such as software development, data analytics, cybersecurity, artificial intelligence, and network administration. The research aims to explore the relationship between students' demographic characteristics, educational backgrounds, and their interest levels in the IT sector. Additionally, the study examines the perceived benefits, career prospects, and social impact associated with the IT industry.

B. Methodology:

Methodology: Data was collected from 676 undergraduate students across XI, XII, BA, B.Com., B.B.A. and B.C.A. faculties using a questionnaire distributed through Google Forms. Random sampling techniques were employed, and 50 students were randomly selected from each class for data analysis. The collected data stored in a CSV file and analysed using Python programming.

Data Collection:

A survey questionnaire developed to collect data from a sample of students. The questionnaire include questions related to students' demographic information, educational backgrounds, awareness of the IT sector, interest levels in different IT domains, and factors influencing their interest.

Sample Selection: The study target students from various educational institutions, including colleges, and technical schools. A stratified sampling approach may be employed to ensure representation across different educational levels, disciplines, and geographic locations.

Data Analysis: The collected data will be subjected to data science techniques and statistical analyses. Descriptive statistics will be used to examine the distribution of interest levels and

demographic characteristics. Correlation analysis will be conducted to identify relationships between interest levels and factors such as educational background, and awareness of the IT sector. Machine learning algorithms may be employed to predict students' interest in specific IT domains based on their demographic information and other variables.

VI. Literature Review:

Data Science Techniques for Analyzing Educational Data: The application of data science techniques in analyzing educational data has gained prominence in recent years. Data mining, machine learning, and statistical analyses have been employed to extract insights and patterns from large-scale educational datasets. These techniques have been used to identify factors influencing students' academic performance, career choices, and interest in specific fields, including the IT sector.

Implications for Education and Industry: The findings from previous research have several implications for educational institutions and industry professionals. First, understanding the factors that influence students' interest in the IT sector can aid in developing targeted interventions and educational programs to attract and retain students in this field. Second, the application of data science techniques in analyzing educational data can provide valuable insights for curriculum development, career guidance.

Data Collection for Research Paper: Analyzing Students' Interest in the Information Technology Sector Using Data Science Techniques. Employ the following survey questionnaire design and data collection methods:

VII. Questionnaire Structure:

A. Introduction: Begin with a brief introduction explaining the purpose of the survey and assuring participants about the confidentiality and anonymity of their responses.

B. Demographic Information: Include questions to gather participants' demographic details, such as age, gender, educational background, and academic level.

C. Interest Levels: Measure participants' interest in the IT sector rating scales. Include questions about overall interest and specific domains within IT (e.g., software development, data analytics, cybersecurity).

D. Factors Influencing Interest: Assess the factors influencing students' interest in the IT sector. Include questions on intrinsic motivations, career prospects, educational experiences, exposure to technology, role models, and other relevant factors.

E. Awareness and Knowledge: Evaluate participants' awareness and knowledge of different IT domains, their applications, and potential career opportunities within the IT sector.

F. Educational Experiences: Gather information on participants' exposure to IT-related

coursework, internships, extracurricular activities, or industry collaborations during their academic journey.

G. Open-Ended Questions: Include a section for participants to provide additional insights, experiences, or suggestions related to their interest in the IT sector.

Online Surveys: Utilize online survey platforms (e.g., Google Forms) to design and distribute the survey questionnaire. Online surveys offer convenience and ease of data collection, allowing participants to respond at their own pace and from various locations. Ensure the survey is mobile-friendly for better accessibility.

VIII. Data Analysis:

Data analysis is a practice in which raw data is ordered and organized so that useful information can be extracted from it [2]. The process of organizing and thinking about data is key to understanding what the data does and does not contain [5]

Types of Data Analysis There are several types of data analysis techniques that exist based on business and technology. The major types of data analysis are:

1. Text Analysis
2. Statistical Analysis
3. Diagnostic Analysis
4. Predictive Analysis
5. Prescriptive Analysis

To analyze students' interest in the Information Technology (IT) sector using data science techniques for the research paper, the following data analysis methods can be employed:

A. Data Preprocessing:

Clean the collected data by handling missing values, outliers, and inconsistencies. Perform data transformations, such as normalization or standardization, to ensure consistency across variables. Check for data quality and address any data issues that may affect the analysis.

B. Descriptive Analysis:

Conduct descriptive analysis to gain an overview of the data and identify key trends and patterns.

Calculate summary statistics (mean, median, standard deviation) to understand the central tendencies and variations in students' interest levels.

C. Create visualizations:

Such as Ogive chart, to present the distribution of interest levels and explore relationships with other variables.

D. Inferential Analysis:

Perform inferential analysis to examine relationships between different variables and their impact on students' interest in the IT sector.

E. Conduct correlation analysis:

To determine the strength and direction of associations between variables, such as demographic characteristics, educational experiences, and interest levels.

F. Predictive modeling:

Utilize predictive modelling techniques, such as logistic regression, decision trees, or random forests, to predict students' interest in the IT sector based on their demographic characteristics and other relevant variables.

Split the data into training and test sets and evaluate the performance of the predictive models using appropriate metrics (e.g., accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve).

Interpret the results of the predictive models to understand the factors that significantly influence students' interest and identify the most influential variables.

Matplotlib: Visualization with Python [8]

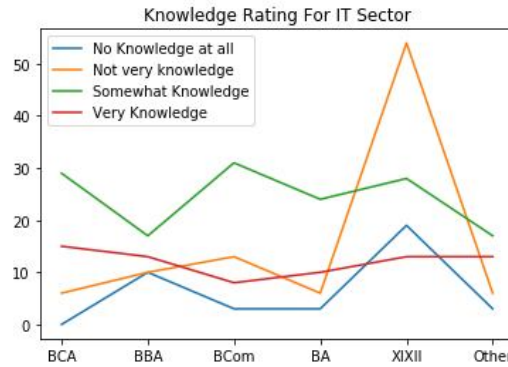
Data visualizations are an important method of sharing your data with others. Python has many 3rd party packages that do data visualizations. In fact, there are so many that it can be somewhat overwhelming. One of the oldest and most popular is Matplotlib. Matplotlib is known for creating static, animated, and interactive visualizations in Python.

The Matplotlib library is used by the Panda library to produce visuals. Using script, Matplotlib is used to make 2D graphs. ^[9]

XI. Table and Chart

1.1 Knowledge Rating IT Sector

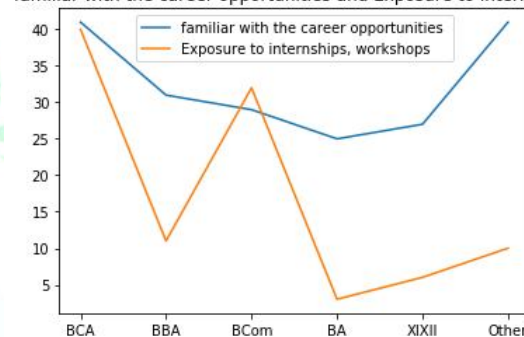
class	No Knowledge at all	Not very knowledge	Somewhat Knowledge	Very Knowledge
Bca	0	6	29	15
Bba	3	8	32	7
bcom	0	11	31	8
Ba	3	13	24	10
XI/XII	10	10	17	13
Other	3	6	28	13



1.2 Familiar with the career opportunities available in the IT sector

Class	Yes	No
BCA	41	9
BBA	31	19
Bcom	29	21
BA	25	25
XI/XII	27	23
Other	41	9

familiar with the career opportunities and Exposure to internships



1.3 Currently pursuing or planning to pursue a degree or specialization in the IT sector?

Class	No, not planning to pursue	Yes, currently pursuing	Yes, planning to pursue
Bca	0	50	0
BBA	9	11	30
Bcom	19	8	23
BA	21	8	21
XI/XII	22	7	21
Other	2	27	19

Not Planning , currently pursuing or planning to pursue a degree or specialization in the IT sector

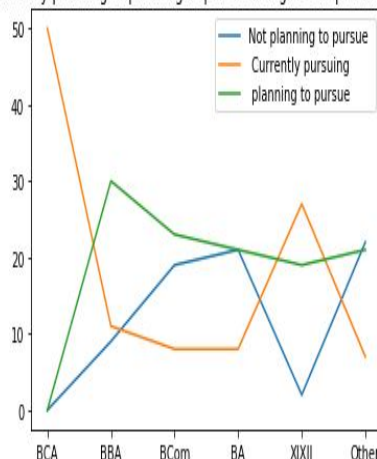


Table Interpretation:

From table 1.1 we see that the students from BCA, XI/XII and other (mostly from Engg. Field) are very knowledgeable in the IT Sector. BCA, B.Com. and other students are wellFamiliar with the career opportunities available in the IT sector. 100 % bca studens are currently pursuing specialization in It sector and BBA. (60%), B.Com.. (46%), BA (42%) . XI/XII (42%) and Other (38%) are planning to pursue the some courses related to IT

X. Conclusion:

In conclusion, this research paper emphasizes the significance of Artificial Intelligence (AI), Data Science, Cloud Computing, and Cybersecurity as advanced and promising fields for career opportunities. The findings highlight the need to raise awareness among students about these domains and provide them with hands-on experiences through internships and workshops.

The rapid advancements in technology have reshaped various industries, and AI, Data Science, Cloud Computing, and Cybersecurity have emerged as key drivers of innovation and digital transformation. It is essential for students to be aware of these fields to align their educational and career paths with future-focused opportunities.

By creating awareness about AI, Data Science, Cloud Computing, and Cybersecurity, educational institutions and stakeholders can empower students to make informed career choices. Students who are familiar with these fields can explore diverse and rewarding career paths, contributing their skills to tackle global challenges and shape the future.

Providing students with hands-on experiences through internships and workshops is vital to bridge the gap between theoretical knowledge and practical application. Internships offer opportunities to work on real-world projects, build industry connections, and develop sought-after skills. Workshops, on the other hand, provide structured learning environments, enabling students to acquire technical expertise and gain confidence in their abilities.

As AI, Data Science, Cloud Computing, and Cybersecurity become integral to various

industries, students equipped with hands-on experience and relevant skills will be better positioned to meet the demands of the job market. Through internships and workshops, students can explore their interests, build industry-relevant portfolios, and be well-prepared to embark on successful and impactful careers in these advanced fields.

In conclusion, raising awareness and providing hands-on experiences in AI, Data Science, Cloud Computing, and Cyber security are crucial steps towards preparing a skilled and future-ready workforce, fostering innovation, and ensuring the sustained growth and progress of the global technology landscape. By empowering students with knowledge and practical experience, we can nurture the next generation of professionals who will drive transformative advancements and shape a digitally empowered world.

These findings highlight the importance of promoting IT career awareness and providing students with opportunities to gain hands-on experience through internships and workshops. Additionally, addressing the lack of relevant education and training can help prepare students for the diverse and rapidly evolving IT industry.

REFERENCES:

- [1] TIOBE Software Index (2011). "TIOBE Programming Community Index Python".^[1]
- [2] "Programming Language Trends - O'Reilly Radar". Radar.oreilly.com. 2 August 2006.^[2]
- [3] "The RedMonk Programming Language Rankings: January 2011 – tecosystems". Redmonk.com.^[3]
- [4] Summerfield, Mark. Rapid GUI Programming with Python and Qt^[4]
- [5] Mohaiminul Islam. Data Analysis: Types, Process, Methods, Techniques and Tools. International Journal on Data Science and Technology. Vol. 6, No. 1, 2020, pp. 10-15. doi: 10.11648/j.ijdst.20200601.12^[5]
- [6] <https://www.ijert.org/research/data-analysis-using-python-IJERTV10IS070241.pdf>
- [7] <https://www.ijres.org/papers/Volume-9/Issue-11/Ser-2/G09113653.pdf>^[7]
- [8] <https://ieeexplore.ieee.org/document/8701341>
- [9] <https://actascientific.com/ASCS/pdf/ASCS-04-0236.pdf>
- [10] <https://arxiv.org/ftp/arxiv/papers/1911/1911.08271.pdf>
- [11] <https://amity.edu/UserFiles/ajjem/1168-converted.pdf>