A STUDY ON DATA ANALYSIS USING PYTHON ON DLK TECHNOLOGIES PRIVATE LIMITED IT SOLUTIONS

A Summer Internship Report submitted to the **SRM INSTITUTE OF SCIENCE AND TECHNOLOGY** in partial fulfilment of the requirements for the award of the Degree of

MASTER OF BUSINESS ADMINISTRATION

Submitted by

SHAKTHI VELU A (RA2152008010019)

Under the guidance of

Dr. A R SHANMUGAPRIYA (Faculty Guide)

COLLEGE OF MANAGEMENT

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR- 603 203

JULY 2022

DLK

Career Development

CERTIFICATE OF COMPLETION

This is to certify that

MR. SAKTHI VELU. A (REG NO: RA2152008010019)

has Successfully Completed

Internship / ~~Inplant training~~ in DATA SCIENCE from 03/06/22 to 16/07/22

During the training period the performance of the trainee was found to be excellent.

Congratulations on your hard work! keep shining!

Jayalakshmi S
Vice president

BONAFIDE CERTIFICATE

This is to certify that the Summer Training Report entitled **"A study on data analysis using python on DLK TECHNOLOGIES PRIVATE LTD IT SOLUTIONS"**, in partial fulfilment of the requirements for the award of the Degree of **Master of Business Administration** is a record of original training undergone by **SHAKTHIVELU A (RA2152008010019)** during the year **2022-2023** of his study in the College of Management, **SRM IST,** Kattankulathur under my supervision and the report has not formed the basis for the award of any Degree/Fellowship or other similar title to any candidate of any University.

Place:

Signature of Guide

Date:

Dr A R SHANMUGAPRIYA
Assistant professor
College of Management
SRM IST, Kattankulathur

Submitted to the College of Management, SRM IST, Kattankulathur for the examination held on_____

PROGRAM CO-ORDINATORCHAIRPERSON-EDUCATION PROGRAM

DEAN-COM

INTERNAL EXAMINER **EXTERNAL EXAMINER**

DECLARATION

I, **SHAKTHI VELU A,** hereby declare that the Summer Training Report, entitled **"An organisational study on "DLK TECHNOLOGIES PRIVATE LIMITED IT SOLUTIONS"**, submitted to the **SRM IST** in partial fulfilment of the requirements for the award of the Degree of Master of Business Administration is a record of original training undergone by me during the period **<Date>** under the supervision and guidance of **Dr A R SHANMUGAPRIYA**, SRM IST, Kattankulathur and it has not formed the basis for the award of any Degree/Fellowship or other similar title to any candidate of any University.

Place: **Signature of the Student**

Date:

ACKNOWLEDGEMENT

First and foremost, I offer my sincerest gratitude to our **Chancellor**, SRM University, for his academic support and the facilities provided to carry out the project work at the Institute. His wide vision and concern for students have been inspirational.

I express my heartfelt thanks to our **Dean,** College of Management, SRM IST, and Kattankulathur who provided all facilities for carrying out this project.

I take this opportunity to express my profound gratitude and deep regards to my guide **DR. A R SHANMUGAPRIYA** for the exemplary guidance, monitoring and constant encouragement throughout the course of this project.

I also take this opportunity to express a deep sense of gratitude to **Ms Logeswari** for his/her cordial support, valuable information and guidance, which helped me in completing this task through various stages. I owe my wholehearted thanks and appreciation to the entire staff of the company for their cooperation and assistance during the course of my project.

I thank God Almighty for showering his perennial blessing on me for giving me the courage to pursue this project work successfully. I owe a lot to my parents, who encouraged and helped me at every stage of my personal and academic life, and longed to see this achievement come true.

SHAKTHI VELU A

(RA2152008010019)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABBRIEVIATED FORM	EXPANDED FORM
AI	Artificial Intelligence
BPM	Business Process Management
CGPA	Cumulative Grade Point Average
CRM	Customer Relationship Management
CSS	Cascading Style Sheets
EDA	Exploratory Data Analysis
GDP	Gross Domestic Product
GRE	Graduate Record Examination
HTML	Hypertext Markup Language
IDA	International Development Association
IT	Information Technology
LOR	Letter Of Recommendation
PHP	Hypertext Pre-Processor
PR	Public Relations
SEO	Search Engine Optimization
SOP	Standard Operating Protocol
TOEFL	Test Of English as a Foreign Language
YoY	Year-Over-Year

INTRODUCTION:

This post presents a reference implementation of an employee turnover analysis project that is built by using Python's library. In this article, we introduce Logistic Regression, Random Forest, and Support Vector Machine. We also measure the accuracy of models that are built by using Machine Learning, and we assess directions for further development. And we will do all of the above in Python. Let's get started!

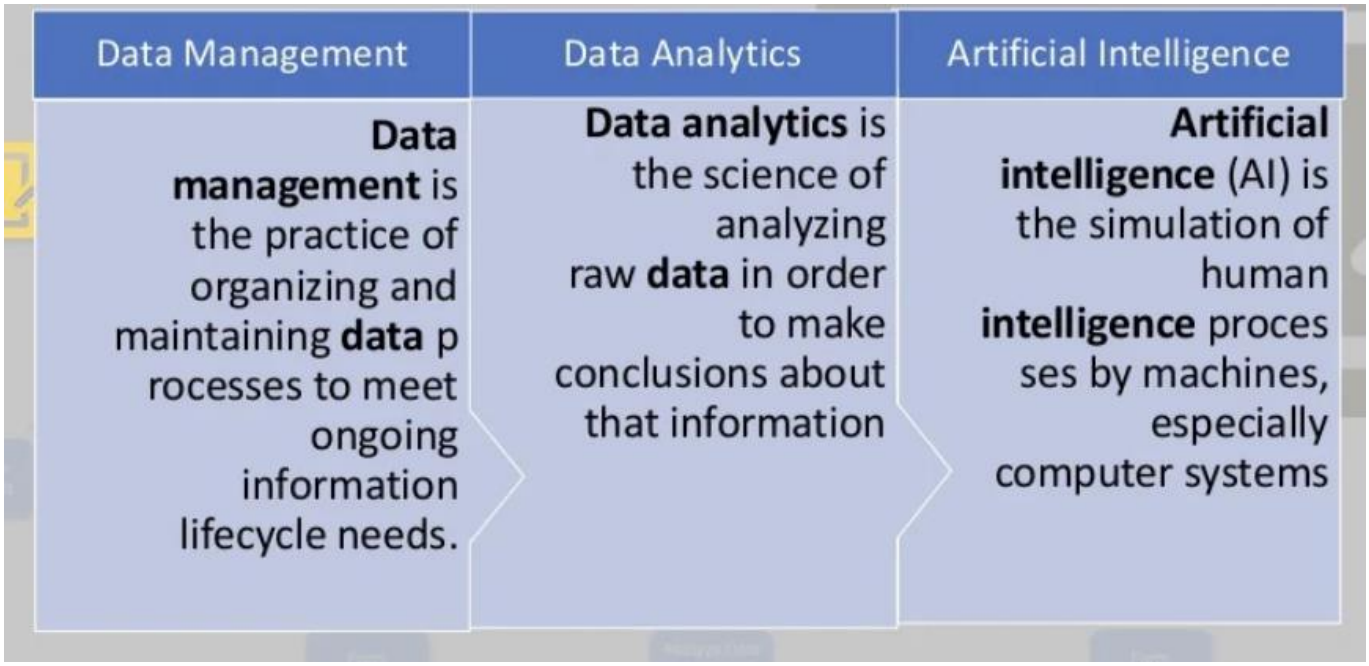
Python is open source, interpreted, high level language and provides great approach for object-oriented programming. It is one of the best language used by data scientist for various data science projects/application. Python provide great functionality to deal with mathematics, statistics and scientific function. It provides great libraries to deals with data science application.

One of the main reasons why Python is widely used in the scientific and research communities is because of its ease of use and simple syntax which makes it easy to adapt for people who do not have an engineering background. It is also more suited for quick prototyping.

No matter who you are or what you're interested in, it's difficult to think about going to college without thinking about what comes after. Your time in higher education is an important period for you to grow, meet new people and experience life from a different perspective, but ultimately, it's about preparing you for your career path and your life after school is done.

As you get ready to enter college, you should get used to hearing a word that's going to start coming up a lot: internships. Internships are becoming more and more important as a way to complement your education and secure your future career. They give you a valuable glimpse into what it's like to work for a living, lend insights into how your chosen field works and what it takes to succeed, and they can show a potential employer you are serious about wanting to be a valuable, successful employee.

Data science is an essential part of many industries today, given the massive amounts of Data that are produced, and is one of the most debated topics in IT circles. Its popularity has grown over the years, and companies have started implementing data science techniques to grow their business and increase customer satisfaction. In this article, we'll learn what data science is, and how you can become a data scientist. Data science is the domain of study that deals with vast volumes of data using modern tools and techniques to find unseen patterns, derive meaningful information, and make business decisions. Data science uses complex machine algorithms to build predictive models.



Objectives of the study:

The main objective of the study is to,

This analysis is part of a learning framework for postgraduate students, with the primary goal of understanding how research is conducted and the different processes involved. This research is focused on real results from Kaggle, and it identifies the primary cause of employee turnover in a different industry. Any organisation's success is highly dependent on its workforce; employees are seen as the company's backbone. The research aids in determining the general interest and attitude of workers against their job and business. It may be used to assess the overall success of the company in terms of employee loyalty and to develop unique retention techniques.

- Satisfaction level (0–1), last evaluation (Time since last evaluation in years)
- number projects (Number of projects completed while at work), average monthly hours (Average monthly hours at workplace)
- time spend company (Time spent at the company in years), Work accident (Whether the employee had a workplace accident)
- left (Whether the employee left the workplace or not (1 or 0)), promotion last 5years (Whether the employee was promoted in the last five years)
- sales (Department in which they work for), salary (Relative level of salary)
- To understand why Python is a useful scripting language for developers.
- Collecting data & Processing data

- To research the concepts and organisational structure involved there
- Analysing (data) and/or applying machine learning (to data)
- To study about the implementation of core management concepts in the industry environment.
- The key objective of Data Science is to extract valuable information for use in strategic decision making, product development, trend analysis, and forecasting.

Python's expansive library of open source data analysis tools, web frameworks, and testing instruments make its ecosystem one of the largest out of any programming community. Python is an accessible language for new programmers because The community provides many introductory resources Programming languages have unique ecosystems, cultures and philosophies built around them. You will find friction with a community and difficulty in learning if your approach to programming varies from the philosophy of the programming language you've selected.

The company's working methodology and learning more about the various departments are the secondary goals. We can better comprehend the functions of each position inside the organisation by learning about other departments. Each division is a component of the whole. When everything is put together properly, the result is a profitable company. Finding out more about the company we work for is always advantageous, regardless of whether we have plans to launch our own business, are considering a change in careers, or are simply inquisitive.

Capture:

Data Acquisition, Data Entry, Signal Reception, Data Extraction. This stage involves gathering raw structured and unstructured data.

Maintain:

Data Cleansing, Data warehousing, Data Staging, Data Processing, Data Architecture. This stage covers taking the raw data and putting it in a form that can be used.

Process:

Clustering/Classification, Data Modeling, Data Summarization. Data scientists take the prepared data and examine its patterns, ranges, and biases to determine how useful it will be in predictive analysis.

Analyse:

Exploratory/Confirmatory, Predictive Analysis, Regression, Text Mining, Qualitative Analysis. Here is the real meat of the lifecycle. This stage involves performing the various analyses on the data.

Communicate:

Data Reporting, Data visualisation business, Decision Making. In this final step, analysts prepare the analyses in easily readable forms such as charts, graphs, and reports.

Importance of the study:

In computing, data is information that has been translated into a form that efficient for movement or processing.Relative Today's computers and transmission media, data is information converted into binary digital form. It is acceptable for data to be used as a singular subject or a plural subject. Raw data is a term used to describe data in its most basic digital format.

Employee Turnover Using Machine Learning With Python project is a desktop application
which is developed in Python platform.

Data science is significant in business for a number of reasons. Enterprises may monitor, track, and record performance measures with the use of data science, which enables better decision-making across the board. Companies can use trend analysis to inform important decisions that will improve consumer engagement, boost organisational performance, and boost profitability. Data science models can replicate several operations using data that is already available. As a result, businesses can plan their strategy for achieving the greatest results. By fusing previously collected data with new information to produce insightful analyses, data science aids enterprises in identifying and defining target audiences. In order to find individuals that best meet their company's demands, data scientists use a combination of data elements.

Data scientists use Python and R for data preparation and statistical analysis. Compared to R, Python is used for general purpose, more readable, simpler, and offers more flexibility while learning. Moreover, Python is used in several verticals other than Data Science and offers you various applications.

From this there will be many benefits for the person working, the employer and for the long term too.

BENEFITS OF STUDY

Apply your theory:

Internships offer students the chance to put what they are learning into action, in a real-world environment. This helps you better understand the theories and strategies you have been reading about, cementing the learning process and giving you greater focus.

Get a feel for the work environment:

For students who are exploring their career options, internships are great! By joining a team, you will have a much better understanding of what it's like working at a particular company and get a clearer idea of the industry itself.
This knowledge will help you in your job hunting in the future, giving you an better idea of the types of jobs you want – and perhaps more importantly – the types of job you don't want.

Boost your confidence

Of course, taking on an internship helps you learn about the work environment, but it also helps you learn about yourself.
You will have a much clearer idea of your own strengths, weaknesses, likes and dislikes. Most importantly, knowing that you have hands-on experience will give you far more confidence when it comes to job seeking and interviews.

Build networks

As Porter Gale wrote, "your network is your net worth" and internships offer students great networking opportunities. You'll meet colleagues and team members, take part in meetings and get to know new people in a professional environment.
If you distinguish yourself during your internship, you can make life-long connections who can help you find positions, meet clients, or even make recommendations.

Increase your motivation

There's a big difference writing an academic paper and writing a real report, with real world consequences. Knowing that you are contributing to the success of a live project and that people are counting on you to do a good job, gives you additional motivation.
When it comes to classwork, you will also reinforce why you are working towards a qualification – giving you that extra push to study hard.

Improve your CV

Students who put themselves forward for an internship show that they are willing to take responsibility, work hard, want to learn, and are interested in getting experience. These are all qualities that hiring managers are interested in and this helps you differentiate yourself in a competitive jobs market.
No matter how successful you were in your internship, you can hold your head up high and explain what you learned and what responsibilities you had.

Getting a job directly

Work hard and take initiative and you may even be offered a position at the company after your internship and studies are completed. If a manager is confident in your abilities to do a job and knows who you are, you are in with a far greater chance of getting a role than you would otherwise.
Of course this is never guaranteed, but internships certainly help you get your foot in the door.

Exposure to their chosen field:

Up to 80% of college students switch majors at some time during their undergraduate studies. Many people find that after learning more about a topic, it may not end up being what they had hoped it would be when they first started their college careers. Before you fully commit to a career, taking part in internships can be a terrific opportunity to sample what it's like to work there professionally. "Internships have other benefits, such as the chance to 'test drive' a career,'" says Alexander Lowry, professor of finance at Gordon College. "They not only help get a foot in the door with a potential employer and look good on a resume. Christopher K. Lee, career consultant and owner of Purpose Redeemed, says that internships give students a low-pressure way to enter their sector. They can choose whether to continue their pursuit based on these encounters.

Work experience:

One of the greatest frustrations college graduates who struggle to find employment report is the problems that arise when the only open positions require work experience; they've not yet had a chance to acquire. "Fresh out of college, us rarely have anything significant to say in a job interview. No one asks about our accounting class or what grade us got on a group project," explains Kaitlyn Trabuco, founder of Educents by CoLearn, a parenting marketplace for educational resources. "The best real work experiences us can have been in our

internship.” Our experts are united on this point. “An internship provides the work experience that helps students put their education into practice, develop their leadership skills and give them a competitive advantage as they pursue a permanent position,” Lowry says.

Employers can assess emerging talent by: When determining which applicants will be the best fit for a position, hiring managers always assume some level of risk.

Most job applicants "put their best foot forward on paper," according to Lowry. "An internship helps a hiring manager assess a candidate's suitability for the position at hand." By identifying young talent and offering internship opportunities, businesses have the opportunity to conduct an extended interview that goes far beyond normal interviewing techniques.

Young professionals may learn about brands from employers by: Internships can be used by businesses to introduce the next generation of industry professionals to the distinctive qualities, offerings, and requirements of their brand. They can also be used to get a thorough assessment of the soon-to-be graduates.

Because many internships are interwoven with academic programmes, Lee explains the benefits of exposing students who are about to enter the workforce to the company's brand. Employers use these collaborations, in his words, "to advise educators about current requirements and in-demand skills in the workforce."

According to Trabuco, employers may gain from students' new perspectives because they typically offer fresh perspectives and unrivalled excitement. He continues, "As an employer, I found that some of the best, most unique ideas were coming from the interns." One intern claimed, "But the interns were the ones who were prepared to take on the dirty job and pick up the responsibilities others didn't necessarily want to complete," that some of the best achievements came from their initiatives.

Job experience

Job listings often describe requirements such as education and minimum job experience. If you are entering the workforce for the first time after college or a vocational training program, you likely do not yet have the required experience. An internship is a perfect way to fill that gap. You can gain valuable exposure to how a business in your chosen field runs, take part in meetings and perform assigned tasks in a real-world setting.

Research experience

In scientific fields, for example, an internship might consist of assisting with research in a lab. You can test the research skills you learned in your college courses in a practical setting and meaningfully contribute to the important research the lab does. Some permanent research jobs require this kind of post-education training. Internship opportunities can help you decide what kind of lab work you would like to do.

Access to a variety of tasks and departments

While some internships, like lab research, might assign you to a specific task every day, other offices might enable exposure to people in various departments and jobs. You may help senior management, sit in on meetings, complete small tasks for projects or observe the daily functions of the office. If your field of study was general, like a B.A. in English, for example, an internship at a publishing company where you see how each department works could help you decide what kind of job in the publishing industry you might like to pursue.

Mentorship

The most valuable mentor relationships often start with a personal connection that enhances the experience for both the mentor and the mentee. Working as an intern may allow you to meet a potential mentor naturally and establish a relationship that helps guide your career path.

In person, you have the chance to reveal your personality and show deference and respect while seeking guidance and input. An internship can create a situation where your

relationship develops organically, even if you end up working elsewhere in a full-time position.

Help guide career goals

During your education, you may study a variety of subjects to determine your interests. For some, a college degree helps you determine exactly what kind of career you want to have. For others, studying offers a diverse set of experiences that presents multiple career options. An internship can give you job experience by introducing you to daily rigors and tasks without a firm commitment to that precise career track. It can help you decide if certain jobs fit your personality and your talents and allow you to meet people who can give you career advice and guidance.

Scope for the study:

The paper emphasises the value for sales organisations to understand their turnover rates and the factors that contribute to them. On a daily basis, salespeople communicate directly with consumers. Removing work dissatisfied is considered necessary for more happy customers. Following that, the survey's findings are discussed. The survey recognises certain factors, such as politics, position uncertainty, and supervisory problems, that have a significant effect on sales company's attrition rate.

The consequences of these factors for workers in various locations (with differing problems), tenure levels, and grades are discussed. This analysis, however, has a lot of weaknesses. The influence of these variables can be assessed across genders, educational backgrounds, skill levels, and performance levels. The results of an in-depth analysis can be generalised in the context of India's sales industry. This information can be useful for companies trying to figure out what factors affect employee turnover among their salespeople

Following are some of the popular data science careers you can crack with an advanced degree:

- Business Intelligence Developer
- Data Architect
- Applications Architect
- Infrastructure Architect
- Enterprise Architect
- Data Analyst
- Data Scientist
- Data Engineer
- Machine Learning Scientist
- Machine Learning Engineer
- Statistician

A 2018 Kaggle research found that Python was used for data analysis by 83% of data professionals. The highest-paid professionals are data scientists, who should ideally be masters of all crafts with programming expertise, industry experience, and understanding of mathematics and statistics.

The customer service and product marketing industries have experienced a significant paradigm change. The old methods are being replaced by chatbots, virtual reality, and augmented reality.

Block chain technology is on the rise where all the ledgers and transactions are recorded leaving no scope for tampering. The stream which is limited to the finance sector today is likely to get exploited in healthcare, banking and insurance sectors.

A colossal amount of data is being dumped into storage every minute collected from online transactions, in-person purchases, social media, and website interactions. However, organizing and analysing this collected data is a challenge for many companies. This is where the Data Science technologies come in to picture where the structured and unstructured data is studied to draw meaningful interpretations and aid in business growth.

Period of the study:

The period of my study was about 5 weeks at DLK Technologies Pvt Ltd from 3rd June 2022 to 16th July 2021.

CHAPTERIZATION:

CHAPTERS	CHAPTER NAME
I	Introduction
II	Industry and company Profile
III	Functions Of Each Department
IV	Findings of the study
V	Conclusion

Table 1.1 Chapterization

CHAPTER I

In this chapter the introduction and the objective of the study is briefly explained. The scope and importance of this is given in this chapter. The overview of my internship is given here.

CHAPTER II

This chapter explains the overview of the Information Technologies Industry, their growth and how it impacts. The overview of DLK Technologies Pvt Ltd is given and the company profile is explained

CHAPTER III

This chapter gives an overview of the functionality of each department and in-depth knowledge of the company's organizational structure and the work done in the data science department

CHAPTER IV

This chapter the findings of the study will be explained with the data collected.

CHAPTER VI

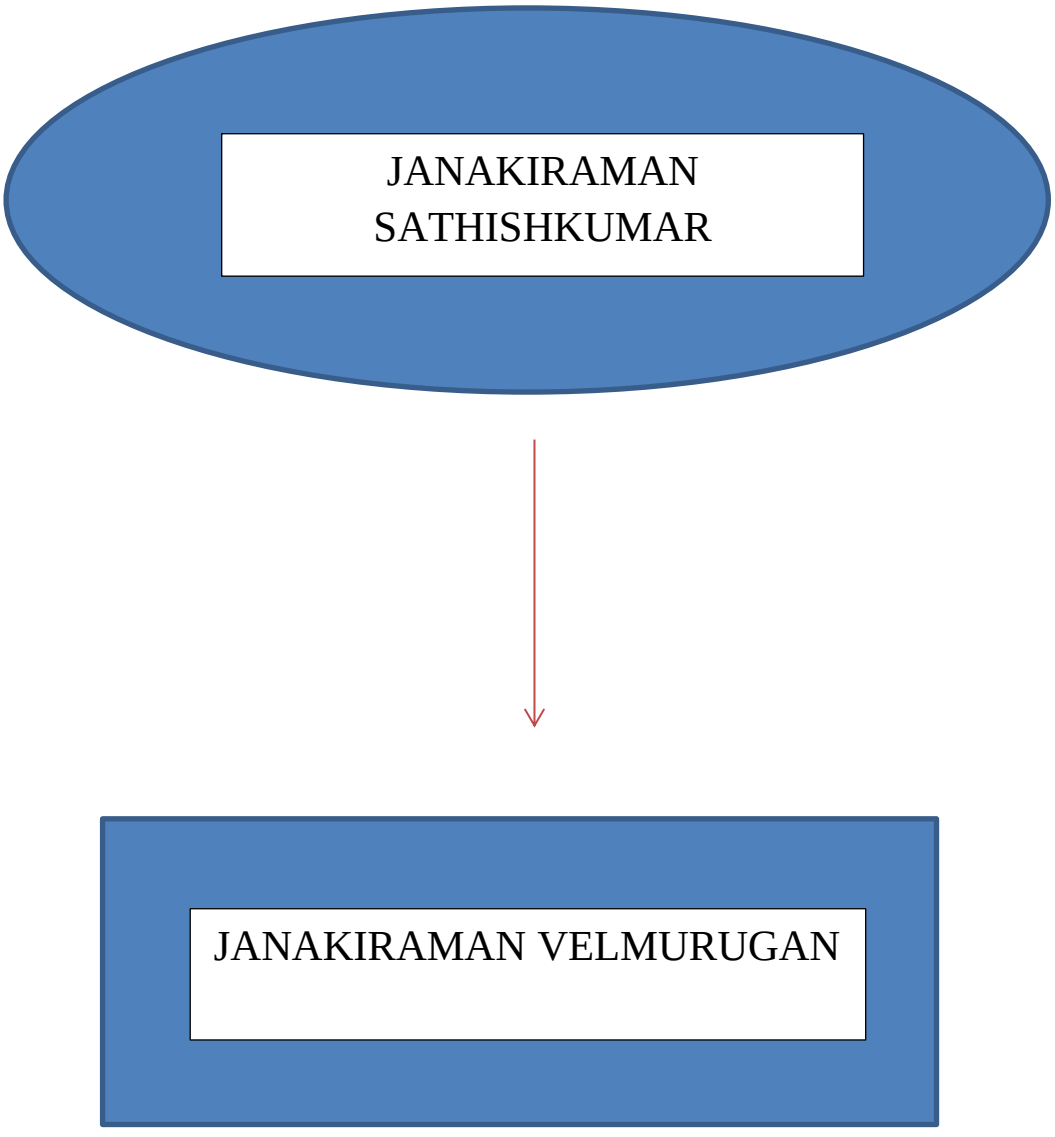
This chapter gives the conclusion and reference of the study.

Structure of the company:

There are two directors for the firm, but no important managerial figures are known.

The two directors that have been on the board the longest are Janakiraman Sathishkumar and Janakiraman Velmurugan, both of whom were appointed on May 5, 2017. For more than 5 years, they have served on the board.

With seats at a total of 1 firm, Janakiraman Sathishkumar has the most additional directorships. The directors of the company have no connections to any other businesses overall.

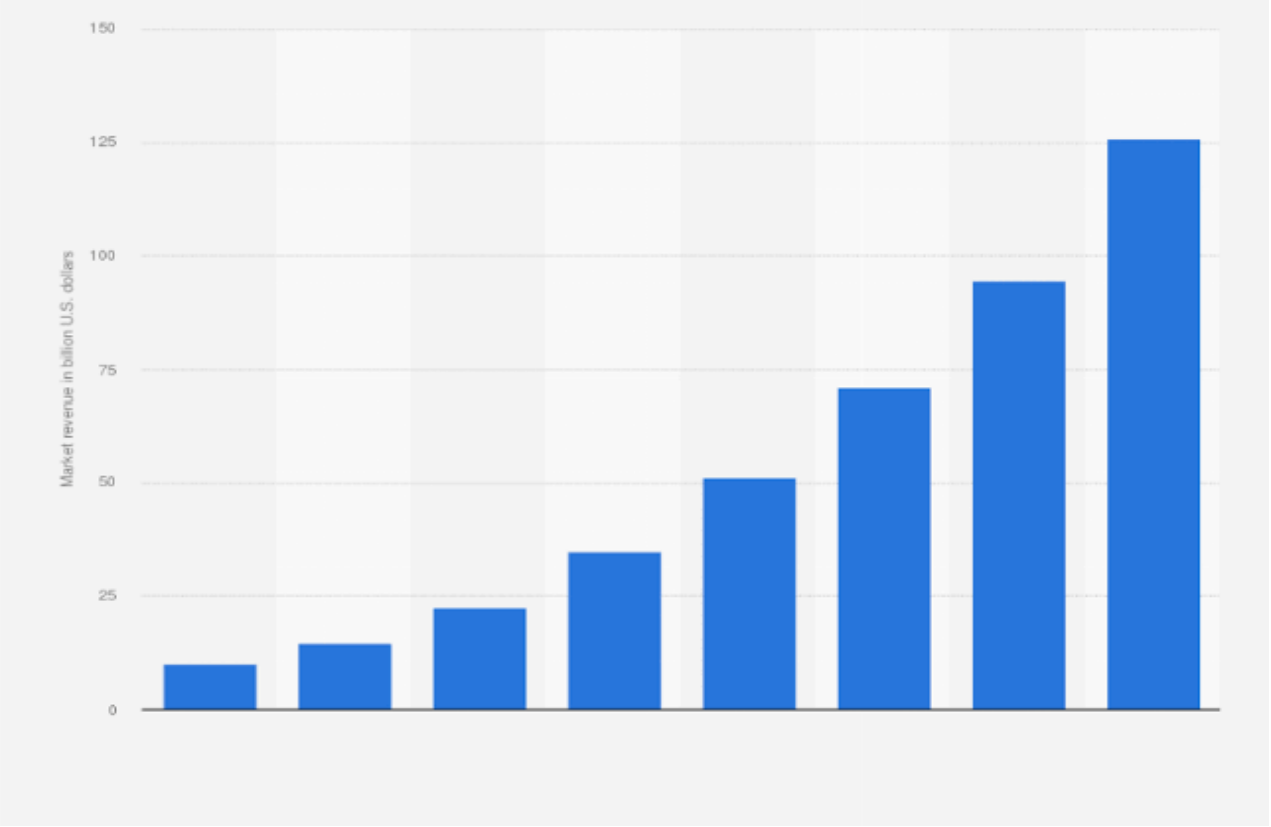


INDUSTRY AND COMPANY PROFILE

INDUSTRY PROFILE

India has a sizeable information technology industry that offers consulting, outsourcing, and IT services. The IT industry made up 8% of India's GDP in 2020. It is anticipated that the IT and BPM sector would generate US$194 billion in revenue in FY 2021, an increase of 2.3 percent year over year. The IT industry is anticipated to provide local revenues of US$45 billion and export revenues of US$150 billion in FY 2021. As of March 2021, 4.5 million individuals worldwide work in the IT-BPM industry. Indian IT-BPM is the sector with the highest employee attrition rate. In recent years, the sector has seen a sharp rise in resignations at all levels of hierarchy. The Indian IT industry is infamous for using low-wage labour as a hub for international outsourcing.

As the IT-BPM industry grows, many individuals are concerned that artificial intelligence (AI) could cause significant automation and job losses. The United States receives two thirds of India's exports of IT services.



2.1 India's IT Industry

The exponential growth of India's IT industry over the past 20 years, which has also supported economic prosperity, has altered the world's perception of the country's knowledge and skill base. The Indian government's liberalisation policies, such as the abolition of import duties on technological items and the lowering of trade barriers, have been instrumental in the sector's rapid expansion. Additional government initiatives, such as the development of Software Technology Parks (STP), Export Oriented Units (EOU), Special Economic Zones (SEZ), and Foreign Direct Investment (FDI), have helped this sector secure a leading position in the global IT sector.

In the information-driven 21st century, India is recognised as a knowledge powerhouse and the centre of global appeal. Software, hardware, e-commerce, IT services, and services made possible by IT are all included in the IT sector (ITES). Any company that wishes to increase efficiency, facilitate business operations, and expand profitably in today competitive world needs IT-based services. In addition to boosting the country's economy, information technology has benefited government by making it more capable and approachable. Government information and services are now easier to get and more reasonably priced. The administration and provision of government services, particularly those pertaining to health care, education, consumer rights, and other sectors, have also been made simpler by information technology. Our economy depends on the IT sector to grow at an exponential rate and create millions of jobs. Growth in the IT industry will encourage us to grow at a rate comparable to China in every field and assist us in snatching up the international market. Indians' socioeconomic position will rise as a result of this.

According to the analysis, domestic IT services might see growth increase by 2-4 percent over the next five years as sectors around the world continue to adopt digitalisation quickly in order to recover more quickly from disruptions brought on by COVID. With a significant portion of business coming from other countries and the deployment of Indian expertise around the world, the future belongs to India. The education sector is undergoing transition thanks to the central government. The last year of a degree programme will be handled as an R&D year in the coming years, and the curriculum is being created using the most recent technology innovations, skills, and knowledge. This would undoubtedly increase the worth of our graduates so that they become job producers rather than just job seekers.

COMPANY PROFILE:

INTRODUCTION

HI-FI web development and digital marketing company DLK Technologies Private Limited IT Solutions, with its headquarters in Chennai, India, excels in the areas of creative website design and search engine optimisation (SEO), domain registration, web hosting, e-commerce solutions, graphic design, social media marketing, and other software-related programmes.

2.2 Company logo

DLK Technologies Pvt. Ltd. was established in 2010 by the vivacious CEO of the DLK Group, Mr. Satheesh Kumar, and our headquarters are also in Kumbakonam. We Are Young, Excited, Tech Savvy, and Passionate About Staying Ahead Of Innovation Technology Online. All techniques for technical advancement can be incorporated into our projects. Often, Our Brilliant Projects Will Speak for Us!

We offer successful search engine optimization services as well as creative website building using a variety of CMS platforms all over the world. We work hard to please our customers, and we'll help them as they fundamentally expand their businesses. Today, running a firm requires both powerful smart work and hard work. We'll complete both duties for our wonderful customers! We Provide Energized Digital Solutions To Increase Public Awareness Of Your Goods And Services. Throughout The Entire Variety Of Web And Digital Values, Our Exceptional Clientele Trust Us To Deliver Creative, Interactive Reports For Their Clients.

We take great pleasure in expressing our gratitude for the everyday successes we experience as a result of DLK's professionals in execution planning, commitment, website design, and internet marketing. With our cherished clients, they commonly adopt a "fulfil on their demand in every attitude" stance, which makes us more valuable in our successes. As you engage with us, you will certainly realise that you are in the ideal place at the ideal time.

DLK Technologies provides a broad spectrum of IT services in the business applications space with a focus on a number of verticals, including construction, finance, healthcare, education, and related areas. Our software development and maintenance cycles are completed with the highest possible quality at every level thanks to our mature development processes and the stringent quality control practises we use. We have standardised processes for quality assurance, documentation, and the software development life cycle.

DLK Technologies' dedication to a consistent approach, efficient project management strategies, tried-and-true automated tools, quality assurance, testing, and committed specialists accounts for its success in gratifying customers.

Every customer has unique requirements that must be handled in a number of ways, and DLK Technologies is aware of this. We provide a range of business models for precisely this reason. Every model has been developed to offer clients the greatest number of advantages. Below is a list of the various business models that DLK Technologies offers to its customers. They are able to successfully meet their needs with the aid of these models, which subsequently add value for their key industry sectors.

Its success in ensuring client satisfaction may be attributed to DLK Technologies' commitment to a consistent strategy, effective project management techniques, tried-and-true automated tools, quality assurance, testing, and committed specialists. Our software development and maintenance cycles are completed with the highest possible quality at every level thanks to our well-established development processes and the stringent quality management practises we use. We have standardised methods for the life cycle of software development, quality assurance, and documentation.

DLK Technologies provides its software consultants to carry out onsite services. These resources, as a part of the onsite project team would report to Project Manager responsible for the project. The onsite model can be a component of the offsite model, where in, DLK Technologies will send its personnel to the client's side for implementation, testing and support of various projects.

VISION

- Their goal is to establish themselves as the world's the most valuable network through the most recent, top-notch performance, and they each have a specialisation in the work they do.
- Integrity and openness

- Open lines of dialogue with clients

- Adherence to industry norms and provision of high-calibre services

MISSION

- Their goal is to increase the business's potential to generate income in a globalised economy by working efficiently and using the clever "more-than-enough" philosophy.
- Whether we are a start-up or an established brand with annual sales in the millions, they have built their whole business strategy on increasing everything a business requires. With DLK, we can be sure that their team is using a solution that is reliable and protected while still being simple to use.

FUNCTIONING OF VARIOUS DEPARTMENTS:

FINANCE DEPARTMENT:

Making financial decisions for both internal and external affairs is a vital responsibility of a company's finance department. The efficiency of a company's finance department and the employees who make up it determines the operations of that department. Specific responsibilities and functions govern how a financial department is structured and operates.

The division in charge of obtaining and overseeing all financial resources on behalf of the corporation is the finance department. In addition to ensuring that operations run smoothly with the least amount of disruption, the department also monitors income and expenses. In addition to the more traditional responsibilities of handling payroll, sales, and costs, the finance department often performs economic analyses to support significant business projects.

A finance department has specific responsibilities to carry out daily. Its primary functions include:

- Accounting
- Examining financial statements and reporting
- Preparing and forecasting budgets
- Managing operations systems.

Here is a summary of financial information of D.L.K TECHNOLOGIES PRIVATE LIMITED for the financial year ending on 31 March, 2021.

- Revenue / turnover of D.L.K TECHNOLOGIES PVT LTD. is Under INR 1cr
- Net worth of the company has increased by 22.65 %
- EBITDA of the company has increased by 27.36 %
- Total assets of the company has increased by 366.63 %
- Liabilities of the company has increased by 393.07

Operating Revenue	Under INR 1 cr
EBITDA	27.36 %
Net worth	22.65 %
Debt/Equity Ratio	8.78
Return on Equity	23.72 %
Total Assets	366.63 %
Fixed Assets	365.25 %
Current Assets	-37.58 %
Current Liabilities	393.07 %
Trade Receivables	-63.84 %
Trade Payables	281.93 %
Current Ratio	0.03

MARKETING DEPARTMENT:

Every successful business is built on marketing. Their marketing division is quite active and posing as DLK's tour guide. Their area on campus is constantly in a frenzy. They are a committed group that creates separate PPTs using brainstorming, creates charts, and overall causes mayhem. But they didn't know that we might have come because of them.

The several tasks that modern global marketing teams are required to carry out in order to engage consumers with their brand:

Define and maintain our brand: Determining precisely what our brand represents, including its values, qualities, and visions, so that we can communicate this to our audience, is one of the marketing department's key roles.

Create marketing strategies: The decisions that our marketing teams make about everything from the pricing of our products and services to the channels our brand should prioritise should be informed by data.

The marketing team will also be in responsible of creating and administering certain campaigns and initiatives as part of our overall strategy, including the resources required, the length of their duration, the milestones that will be attained throughout that duration, and the analysis of the results.

The best marketing teams conduct in-depth research on the characteristics, tendencies, and motivations of our target audience as well as what our competitors are doing in this space. They also have a thorough understanding of their target market.

Produce assets for our marketing channels: The creation of content and assets for our whole marketing network will be a crucial responsibility of the marketing division. A lengthy list includes social media, email marketing, blogs, print, and digital advertisements (and is only getting longer).

Drive traffic to our website: Generating more high-quality leads for their business and nurturing them for as long as it takes to achieve those crucial conversions is one of the marketing team's top priorities.

Coordinate our social media presence: Everyone needs to be on social media nowadays, so an increasingly vital marketing role is managing and monitoring these platforms to keep these up-to-date and protect our brand’s reputation. • Identify and utilise advertising opportunities: Marketing teams should be actively locating opportunities to advertise their brand, be that through digital or print platforms, and communicating with organisations that can afford them that space.

Organise internal communications: As well as connecting customers with our brand, marketing departments are also responsible for keeping employees continuously aware of the organisation's values, goals and priorities.

Act as our media liaison: When our brand is mentioned in the media, either in a positive or negative light, a marketing team member will often be in charge of coordinating with the media and preparing communications on these platforms.

Planning and handling events: When we host an event, seminar or webinar, it will typically be our marketing team that will organise and manage these.

Manage third parties: Particularly for small or mid-sized businesses, marketing departments will need to work closely with outside vendors, be they agencies, PR companies, influencers or freelancers, and ensure they produce work in-line with our brand's values.

DEVELOPING TEAM

It is about the driving forces behind the accomplishment and the heart of DLK. They have been putting together the perfect group of visionary developers who share their customers' ambition to move rapidly, think creatively, and keep a positive attitude. They will take extra measures to create a project team that accurately represents our special beliefs. The very core of our business will be reached by their programmers, who will unleash the full power of our software.

Their crew will have us covered whether we're seeking for a standout talent set or just a helping hand while our significant other is away.

The Development Team is split into functional sections based on the knowledge and abilities of the staff members. Role assignments are created in accordance with the method chosen for the particular project. A project team for medium- and large-scale projects is made up of the following teams:

- Management (project and group leaders)
- Analysis and Design (system analysts and system architects)
- Engineering (developers, database administrators)
- Quality Assurance and Control (reviewers, testers, test designers)

- Design and Usability (art designers, interface designers)
- Infrastructure Support (system administrators who maintain network,
- source control system, defect-tracking system, workspace and build
- environment)
- Technical Writers and Translators

Project teams are created using the components-and-rolls method, in which team leaders and specialists from different function teams are allocated to different project areas. If the component that has to be produced has a limited scope, one team member may combine numerous functions. This idea of setting up project teams with the appropriate skills provides the quickest development with the fewest defects.

SEO Department:

For the post of SEO Specialist, a highly technical understanding of SEO is necessary. This includes expertise in developing and putting into practise content marketing strategies, keeping an eye on search engine algorithms established by search engines, and producing meta information.

Business Development Managers, SEO Managers, SEO Team Leaders, Junior SEOs, Link Builders, Content Writers, Web Designers, and Developers are additional members of the SEO team. Each member is committed to being the best in their area of expertise. Each one is a stand-alone expert in digital marketing and is conscious of how the success of the internet and SEO strategy as a whole is impacted by their work. They attract the best talent. They work hard to hone their abilities. The SEO team members also show their commitment to DLK.

The overall tasks of an SEO Manager would include:

- Website, social media, and content optimization
- Managing website content
- Building and nurturing an online community
- Negotiating backlinks
- SEO strategy
- Campaign organization and planning
- Website monitoring and user analysis

- General digital marketing project management
- Keyword research
- Market analysis

DESIGNING TEAM:

The young and dynamic design team at DLK includes the imaginative and eccentric people. Without using words, they produce bright images that explain everything.

These tasks are all necessary for developing and maintaining an app or website. Below is a list of DLK Technologies' web development teams' usual positions.

Project Manager:

The team's numerous initiatives are given priority by the digital project manager, who also helps to keep everything on track. They oversee strategy, manage the budget, and make ensuring the team is filled with the greatest individuals.

Analyst: The purpose of this function is to act as a middleman between the client, design, and development. He converts the client's requests (requirements) into precise guidelines for how a certain capability should operate. Once the team has supplied these functionalities, functional analysts will test them as well.

Back-end developer: The code created by this programmer manages what is shown online. Java,.NET, and PHP are just a few of the programming languages the back-end developer uses. To build the webpage, these languages are run on the server. We refer to it as "server side" or "back-end" for this reason: When a user accesses a website, the browser will download the page from the server. Additionally, the back-end developer interfaces the website or app with other platforms, such as CRM applications and other websites.

Front-end developer: The front-end developer, who works closely with his back-end counterpart, is responsible for the "frontend" or "client side" of the website. Once the page is displayed in the browser, it still needs to have a nice layout. To instruct the browser on how to present the website visually, the developer creates CSS and HTML. The front-end developer also writes JavaScript, which is increasingly utilised for dynamic applications with features like log-in and other user interaction and for making visual animations.

Architect: The server on which the webpages are hosted is configured by this individual. The architect, in contrast to the back-end developer, focuses on everything needed to make a website rather than just the website itself. They are in charge of making the website's ideal environment.

UX designer/Graphic designer: These are frequently distinct responsibilities, although they can also be merged. In order for visitors to understand how to use and navigate through the website, the UX (user experience) designer is responsible for outlining the website in a rough sketch (a wireframe). The website's visual style is developed by the graphic designer in accordance with the corporate style manual. For each functionality, he also develops a visual design that serves as a blueprint for the front-end developer, who will translate it into code.

Quality tester: This person tests every functionality that the team has made in every possible condition before the website becomes available to regular users. When something is not in order, he reports a bug to the developers. The tester does not simply click around in the application but follows a strict plan and procedure to make sure they will test every possibility

COPYWRITING TEAM

They are regarded as outstanding thinkers at DLK. However, neither the air nor the absence of doors is certainties.

CONCLUSION:

The success of DLK Technologies India Pvt Ltd. in meeting the needs of its clients is a result of the company's dedication to a consistent methodology, efficient project management approaches, tried-and-true automated tools, quality assurance, testing, and committed personnel

RESEARCH METHODOLOGY

4.1 REVIEW OF LITERATURE

A number of meticulous studies on Python in data analysis were conducted. Research has been done on papers published in Journals and International Conferences on various marketing attribution techniques and its applications.

- Aindrila Ghosh, Mona Nashaat, James Miller, Shaikh Quader, and Chad Marston, "A Comprehensive Review of Tools for Exploratory Analysis of Tabular Industrial Datasets," Visual Informatics, Volume 2, Issue 4, December 2018, pp. 235-253.
- John T. Behrens, "Principles and Procedures of Exploratory Data Analysis," Psychological Methods, 1997, Vol. 2, No. 2, pp.131-160.
- Chokey Wangmo, "An Exploratory Study on Bank Lending To SME Sector In Bhutan," International Journal of Scientific & Technology Research, volume 6, issue 11, November 2017, pp. 47-51.
- Matthew Ntow-Gyamfi and Sarah Serwaa Boateng, "Credit Risk and Loan Default among Ghanaian Banks: An Exploratory Study," Management Science Letters, Vol. 3, 2013, pp.753–762.
- X. Francis Jency, V. P. Sumathi, Janani Shiva Sri, "An Exploratory Data Analysis for Loan Prediction Based on Nature of the Clients," International Journal of Recent Technology and Engineering (IJRTE),
- Volume-7 Issue-4S, November 2018, pp.176-179.
- K. Ulaga Priya, S. Pushp, K. Kalaivani, A. Sartiha, "Exploratory Analysis on Prediction of Loan Privilege for Customers using Random Forest," International Journal of Engineering & Technology, Vol. 7, Issue 2.21, 2018, pp. 339-341.
- Bogumil M. Konopka, Felicja Lwow, Magdalena Owczarz, Łukasz Łaczmański, "Exploratory Data Analysis of a Clinical Study Group: Development of a Procedure for Exploring Multidimensional Data," PLOS ONE, [Online] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6107146/pdf/pone.0201950.pdf, August23, 2018, pp. 1-21.

- Introduction to Machine Learning using Python [Online], Available: https://www.geeksforgeeks.org/introduction-machine-learning-using-python/
- Exploratory data analysis – From Wikipedia, the free encyclopaedia [Online], Available: https://en.wikipedia.org/wiki/Exploratory_data_analysis

4.2 RESEARCH METHODOLOGY

The research methodology used here is EDA- Exploratory data analysis with Python using Jupiter Notebook.

4.2.1 EXPLORATORY DATA ANALYSIS

The programming language Python, with its English commands and easy-to-follow syntax, offers an amazingly powerful open-source alternative to traditional techniques and applications. Data analytics allow businesses to understand their efficiency and performance, and ultimately helps the business make more informed decisions. For example, an e-commerce company might be interested in analysing customer attributes in order to display targeted ads for improving sales. Data analysis can be applied to almost any aspect of a business if one understands the tools available to process information. The ecommerce companies are analysing the reviews of customer by using proper visualization method. Exploratory Data Analysis (EDA) is an approach to summarize the data by taking their main characteristics and visualize it with proper representations. EDA focuses more narrowly on checking assumptions required for model fitting and hypothesis testing, and handling missing values and making transformations of variables as needed. EDA encompasses IDA. EDA quickly describes the data sets number of rows/columns, missing data, data types and preview. Clean corrupted data; handle missing data, invalid data types and incorrect values. EDA visualize data distributions; bar charts, histograms, box plots. Calculate and visualize correlations (relationships) between variables; heat map.

4.3 DATA COLLECTION

The dataset is collected from secondary sources. The dataset for this research is obtained from SEO team. The dataset consists of various features related to the work done by DLK. The dataset is stored in the form of a CSV file.

Satisfaction Level	Last Evaluation	Number Project	Average Monthly Hours	Time Spend Company	Work Accident	Quit	Promotion Last 5years	Department	Salary
0.38	0.53	2	157	3	0	1	0	sales	low
0.8	0.86	5	262	6	0	1	0	sales	medium
0.11	0.88	7	272	4	0	1	0	sales	medium
0.72	0.87	5	223	5	0	1	0	sales	low
0.37	0.52	2	159	3	0	1	0	sales	low
0.41	0.5	2	153	3	0	1	0	sales	low
0.1	0.77	6	247	4	0	1	0	sales	low
0.92	0.85	5	259	5	0	1	0	sales	low
0.89	1	5	224	5	0	1	0	sales	low

Table 4.1 Dataset (First 10 rows)

ANALYSIS AND INTERPRETATIONS

5.1 DATA ANALYSIS

Data need to be analysed so as to produce good result. Using the result decision can be taken. For example, recommendation system, ranking of the page, demand fore casting, prediction of purchase of the product. There are some leading companies where the review of the customer plays a great role to analyse the factor which influences the review rating. We have used exploratory data analysis (EDA) where data interpretations can be done in row and column format. We have used python for data analysis. it is object oriented, interpreted and interactive programming language. it is open source with rich sets of libraries like pandas, Matplotlib, seaborn etc.

5.2 EXPLORATORY DATA ANALYSIS (EDA)

Primarily, exploratory data analysis is an approach to see what the data can communicate us away from the formal modelling or hypothesis testing task. EDA helps to analyse the data sets to summarize their statistical characteristics focusing on four key aspects, like, measures of central tendency (comprising of the mean, the mode and the median), measures of spread (comprising of standard deviation and variance), the shape of the distribution and the existence of outliers. In the following paragraphs, we have presented a description of these key aspects of EDA.

Specific statistical functions and techniques us can perform with EDA tools include:

- Clustering and dimension reduction techniques, which help create graphical displays of high-dimensional data containing many variables.
- Univariate visualization of each field in the raw dataset, with summary statistics.
- Bivariate visualizations and summary statistics that allow us to assess the relationship between each variable in the dataset and the target variable you’re looking at.
- Multivariate visualizations, for mapping and understanding interactions between different fields in the data.

- K-means Clustering is a clustering method in unsupervised learning where data points are assigned into K groups, i.e., the number of clusters, based on the distance from each group's centroid. The data points closest to a particular centroid will be clustered under the same category. K-means Clustering is commonly used in market segmentation, pattern recognition, and image compression.
- Predictive models, such as linear regression, use statistics and data to predict outcomes.

5.2.1 IMPORT LIBRARIES

Import essential modules and helper functions from NumPy, Matplotlib, and scikit-learn.

```
In [1]: from __future__ import print_function
        %matplotlib inline
        import os
        import warnings
        import numpy as np
        import matplotlib.pyplot as plt
        import matplotlib.image as image
        import pandas as pd
        import pandas_profiling
        plt.style.use("ggplot")
        warnings.simplefilter("ignore")

In [2]: plt.rcParams['figure.figsize'] = (12,8)
```

Fig. 5.1 Importing essential libraries

5.2.2 EXPLORATORY DATA ANALYSIS

It is the first stage of data analysis. Here we can know about the content of the data set and characteristic of data set. It tells about the size of the data. We can find the missing value of data. We can find the possible relationship among data. Data visualization is done by the use of tabular data and understanding the characteristics.

```
In [3]: hr=pd.read_csv('data/employee_data.csv')
        hr.head()
```

Out[3]:

satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	quit	promotion_last_5years	department	salary
0.38	0.53	2	157	3	0	1	0	sales	low
0.80	0.86	5	262	6	0	1	0	sales	medium
0.11	0.88	7	272	4	0	1	0	sales	medium
0.72	0.87	5	223	5	0	1	0	sales	low
0.37	0.52	2	159	3	0	1	0	sales	low

Fig. 5.2 Loading the dataset

```
In [5]: pd.crosstab(hr.salary,hr.quit).plot(kind='bar')
        plt.title("Turnover Frequency on salary Bracket")
        plt.xlabel('Salary')
        plt.ylabel('Frequency of turnover')
        plt.show()
```

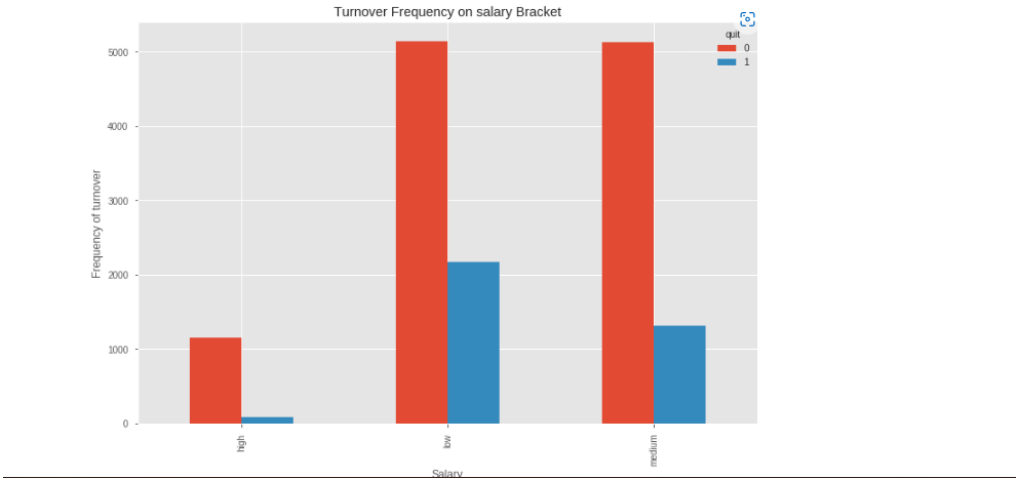


Fig. 5.3 Turnover Frequency on salary Bracket

```
In [6]: pd.crosstab(hr.department,hr.quit).plot(kind='bar')
        plt.title("Turnover Frequency on department")
        plt.xlabel('Department')
        plt.ylabel('Frequency of turnover')
        plt.show()
```

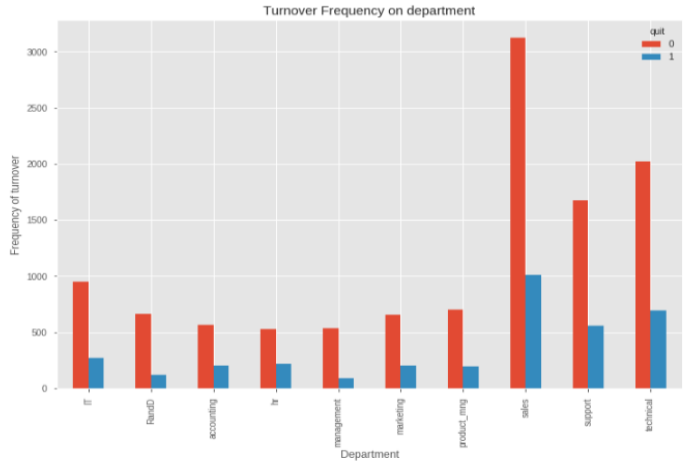


Fig. 5.4 Turnover Frequency on department

5.2.3 ENCODE CATEGORICAL FEATURES

Categorical feature encoding is often a key part of the data science process and can be done in multiple ways leading to different results and a different understanding of input data. A big part of natural language processing is converting text to numbers. Just like that, our algorithms cannot run and process data if that data is not numerical. Therefore, we need to have tools at their disposal to transform colours like red, yellow, and blue into numbers like

1, 2, and 3 for all the backend math to take place. Now that we know what categorical data looks like and have seen some examples, we will examine three common methods to turn our categorical data into numeric data.

The dataset contains two categorical variables: Department and Salary. Create dummy encoded variables for both categorical variables.

```
In [7]: cat_vars=['department','salary']
        for i in cat_vars:
            cat_list=pd.get_dummies(hr[i], prefix=i)
            hr=hr.join(cat_list)

In [8]: hr.head()
```

Out[8]:

	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	quit	promotion_last_5years	department	sa
0	0.38	0.53	2	157	3	0	1	0	sales	
1	0.80	0.86	5	262	6	0	1	0	sales	me
2	0.11	0.88	7	272	4	0	1	0	sales	me
3	0.72	0.87	5	223	5	0	1	0	sales	
4	0.37	0.52	2	159	3	0	1	0	sales	

5 rows × 23 columns

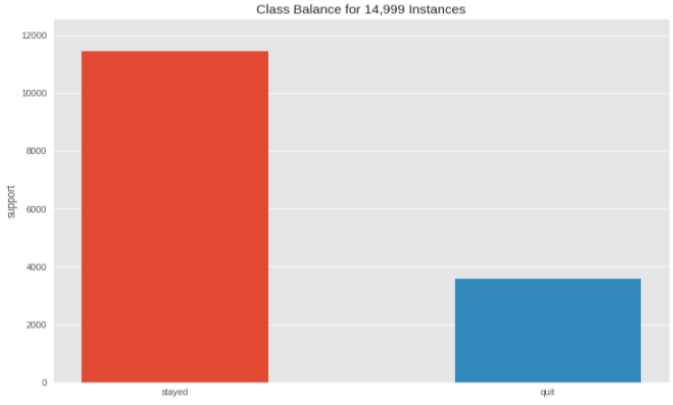
Fig. 5.5 Encode Categorical Features

5.2.4 VISUALIZE CLASS IMBALANCE

Use Yellow brick's Class Balance visualizer to create a frequency plot of both classes. The presence or absence of a class balance problem will inform your sampling strategy while creating training and validation sets.

```
In [10]: from yellowbrick.target import ClassBalance
         plt.style.use("ggplot")
         plt.rcParams['figure.figsize'] = (12,8)

In [12]: visualizer=ClassBalance(labels=['stayed','quit']).fit(hr.quit)
         visualizer.show()
```




```
Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6610924588>
```

Fig. 5.6 Visualize Class Imbalance

5.2.5 CREATE TRAINING AND TEST SETS

Split the data into an 80/20 training/validation split. Use a stratified sampling strategy.

```
In [13]: x=hr.loc[:,hr.columns !='quit']
         y=hr.quit

In [15]: from sklearn.model_selection import train_test_split
         x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=0,test_size=0.2,stratify=y)
```

Fig. 5.7 Create Training and Test Sets

5.2.6 BUILD AN INTERACTIVE DECISION TREE CLASSIFIER

Supervised learning:

The inputs are random variables $X=X1,...,Xp$;

The output is a random variable Y.

Data is a finite set

$\mathbb{L}=\{(xi,yi)|i=0,...,N-1\}$

where $xi \in X=X1\times...\times Xp$ and $yi \in y$ are randomly drawn from PX,Y.

E.g., (xi,yi)=((salary = low, department = sales, ...),quit = 1)

The goal is to find a model $\varphi\mathbb{L}:X\mapsto y$ minimizing

$\text{Err}(\varphi\mathbb{L})=\mathbb{E}X,Y\{L(Y,\varphi\mathbb{L}(X))\}$.

Use the interact function to automatically create UI controls for function arguments. Build and train a decision tree classifier with scikit-learn. Calculate the training and validation accuracies. Display the fitted decision tree graphically.

```
In [16]: from sklearn import tree
         from sklearn.tree import DecisionTreeClassifier
         from sklearn.ensemble import RandomForestClassifier
         from sklearn.metrics import accuracy_score
         from sklearn.tree import export_graphviz # display the tree within a Jupyter notebook
         from IPython.display import SVG
         from graphviz import Source
         from IPython.display import display
         from ipywidgets import interactive, IntSlider, FloatSlider, interact
         import ipywidgets
         from IPython.display import Image
         from subprocess import call
         import matplotlib.image as mpimg

In [32]: @interact
         def plot_tree(crit=['gini','entropy'],
                       split=['best','random'],
                       depth=IntSlider(min=1,max=30,value=2, continuous_update=False),
                       min_split=IntSlider(min=2,max=5,value=2, continuous_update=False),
                       min_leaf=IntSlider(min=1,max=5,value=1, continuous_update=False)):
             estimator=DecisionTreeClassifier(random_state=0,
                                              criterion=crit,
                                              splitter=split,
                                              max_depth=depth,
                                              min_samples_split=min_split,
                                              min_samples_leaf=min_leaf)
             estimator.fit(x_train,y_train)
             print('Decision Tree Training Accuracy: {:.3f}'.format(accuracy_score(y_train,estimator.predict(x_train))))
             print('Decision Tree Testing Accuracy: {:.3f}'.format(accuracy_score(y_test,estimator.predict(x_test))))

             graph=Source(tree.export_graphviz(estimator,out_file=None,
                                               feature_names=x_train.columns,
                                               class_names=['stayed','guit'],
                                               filled=True))
             display(Image(data=graph.pipe(format='png')))
             return estimator

         interactive(children=(Dropdown(description='crit', options=('gini', 'entropy'), value='gini'), Dropdown(descri…
```

Fig. 5.8 Interactive Decision Tree Classifier

5.2.7 BUILD AN INTERACTIVE RANDOM FOREST CLASSIFIER

Use the interact function again to automatically create UI controls for function arguments. To overcome the variance problem associated with decision trees, build and train a random forests classifier with scikit-learn. Calculate the training and validation accuracies. Display a fitted tree graphically.

```
In [34]: @interact
         def plot_tree_rf(crit=['gini','entropy'],
                          bootstrap=['True','False'],
                          depth=IntSlider(min=1,max=30,value=3, continuous_update=False),
                          forests=IntSlider(min=1,max=200,value=100,continuous_update=False),
                          min_split=IntSlider(min=2,max=5,value=2, continuous_update=False),
                          min_leaf=IntSlider(min=1,max=5,value=1, continuous_update=False)):

             estimator=RandomForestClassifier(random_state=1,
                                              criterion=crit,
                                              bootstrap=bootstrap,
                                              n_estimators=forests,
                                              max_depth=depth,
                                              min_samples_split=min_split,
                                              min_samples_leaf=min_leaf,
                                              n_jobs=-1,
                                              verbose=False)
             estimator.fit(x_train,y_train)

             print('Random Forests Training Accuracy: {:.3f}'.format(accuracy_score(y_train,estimator.predict(x_train))))
             print('Random Forests Testing Accuracy: {:.3f}'.format(accuracy_score(y_test,estimator.predict(x_test))))
             num_tree  = estimator.estimators_[0]
             print('\visualizing Tree: ', 0)
             graph=Source(tree.export_graphviz(num_tree,
                                               out_file=None,
                                               feature_names=x_train.columns,
                                               class_names=['stayed','guit'],
                                               filled=True))
             display(Image(data=graph.pipe(format='png')))
             return estimator

         interactive(children=(Dropdown(description='crit', options=('gini', 'entropy'), value='gini'), Dropdown(descri…
```

Fig. 5.9 Interactive Random Forest Classifier

5.2.8 FEATURE IMPORTANCE AND EVALUATION METRICS

Many model forms describe the underlying impact of features relative to each other. Decision Tree models and Random Forest in scikit-learn, feature_importances_ attribute when fitted. Utilize this attribute to rank and plot the features.

```
In [35]: from yellowbrick.model_selection import FeatureImportances
         plt.rcParams['figure.figsize'] = (12,8)
         plt.style.use("ggplot")
```

```
In [36]: rf=RandomForestClassifier(bootstrap='True', class_weight=None, criterion='gini',
                     max_depth=5, max_features='auto', max_leaf_nodes=None,
                     min_impurity_decrease=0.0, min_impurity_split=None,
                     min_samples_leaf=1, min_samples_split=2,
                     min_weight_fraction_leaf=0.0, n_estimators=100, n_jobs=-1,
                     oob_score=False, random_state=1, verbose=False,
                     warm_start=False)


         viz=FeatureImportances(rf)
         viz.fit(x_train,y_train)
         viz.show();
```

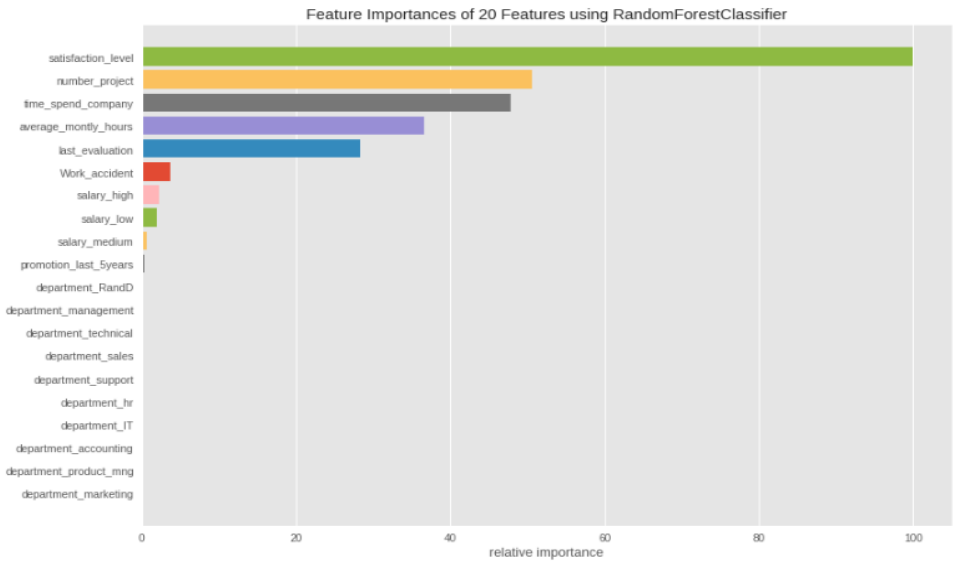


Fig. 5.10 Feature Importance of 20 Features using Radom Forest Classifier

```
In [37]: dt=DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=3,
                     max_features=None, max_leaf_nodes=None,
                     min_impurity_decrease=0.0, min_impurity_split=None,
                     min_samples_leaf=1, min_samples_split=2,
                     min_weight_fraction_leaf=0.0, presort=False, random_state=0,
                     splitter='best')
         viz=FeatureImportances(dt)
         viz.fit(x_train,y_train)
         viz.show();
```

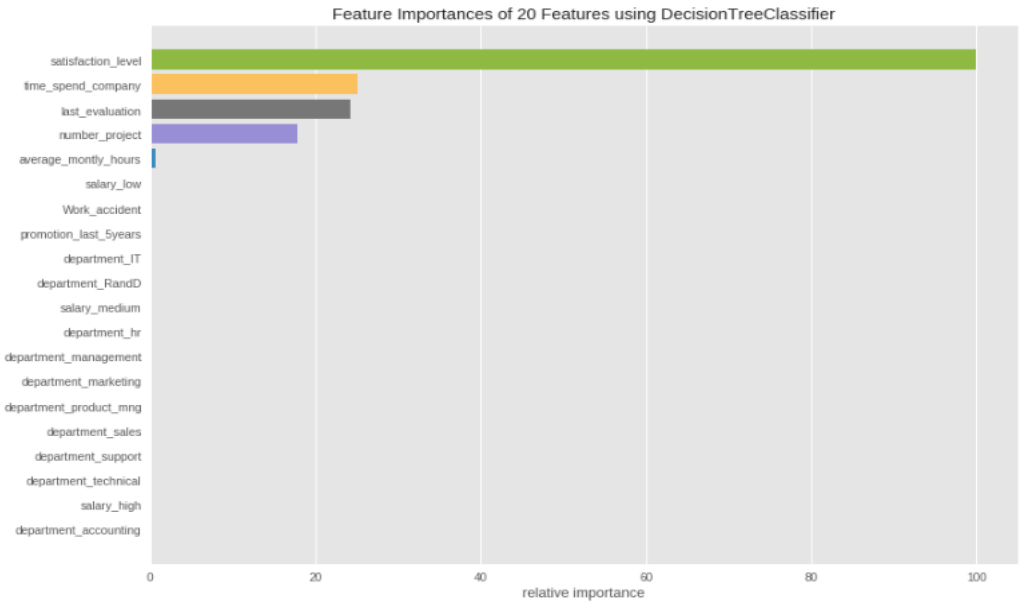
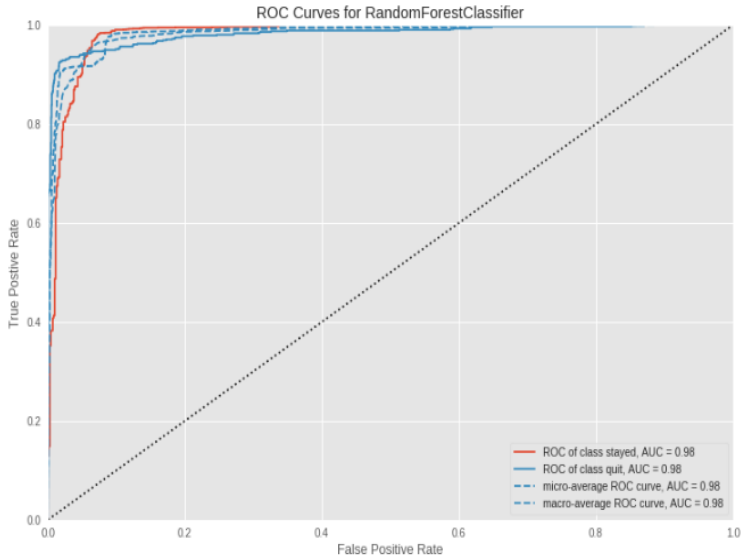


Fig. 5.11 Feature Importance of 20 Features using Decision tree Classifier

```
In [39]: from yellowbrick.classifier import ROCAUC

         visualizer=ROCAUC(rf,classes=['stayed','quit'])
         visualizer.fit(x_train,y_train)
         visualizer.score(x_test,y_test)
         visualizer.poof()
```

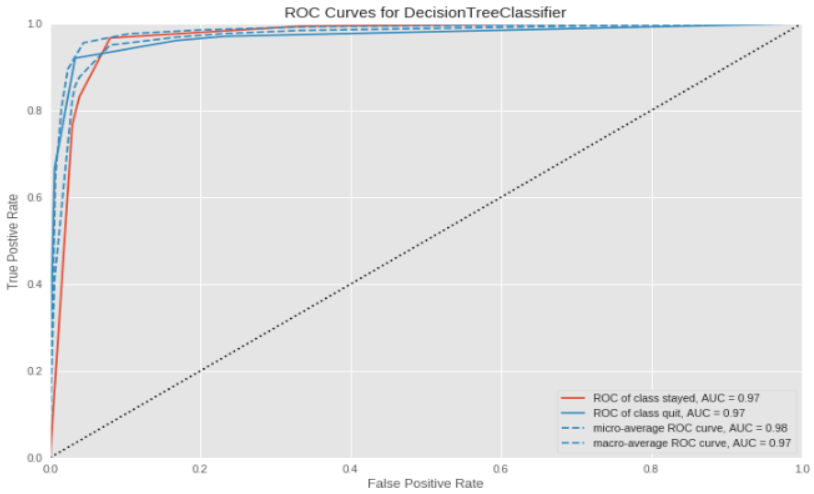



```
Out[39]: <matplotlib.axes._subplots.AxesSubplot at 0x7f66106c6550>
```

Fig. 5.12 ROC Curves for Random Forest Classifier

```
In [40]: from yellowbrick.classifier import ROCAUC

visualizer=ROCAUC(dt,classes=['stayed','quit'])
visualizer.fit(x_train,y_train)
visualizer.score(x_test,y_test)
visualizer.poof()
```




```
Out[40]: <matplotlib.axes._subplots.AxesSubplot at 0x7f6611759780>
```

Fig. 5.13 ROC Curves for Decision tree Classifier

5.2.9 PRESENT RESULT

We can visualise large amount of complex data by the use of chart, graph and tables. Human brain can process information using chart, graphs. It is an easy way to convey the concept. It can identify the area which needs improvement. It can clarify the factor very well.

5.3 INTERPRETATIONS

The following interpretations on the Predict Employee Churn done on the dataset provided using Python.

The most important features using Radom Forest Classifier,

- Satisfaction Level
- Number Project
- Time Spent in Company
- Average Monthly Hour
- Last Evaluation
- Work Accidents

The most important features using Decision tress Classifier,

- Satisfaction Level
- Time Spent in Company
- Last Evaluation
- Number Project
- Average Monthly Hour

From the ROC curves of Radom Forest Classifier,

- ROC of class Stayed, AUC= 0.98
- ROC of class quit, AUC= 0.98
- Micro average ROC curve, AUC= 0.98
- Macro average ROC curve, AUC= 0.98

From the ROC curves of Decision Tree Classifier,

- ROC of class Stayed, AUC= 0.97
- ROC of class quit, AUC= 0.97
- Micro average ROC curve, AUC= 0.98
- Macro average ROC curve, AUC= 0.97

CONCLUSION:

Employee attrition is a major problem that the Indian IT industry has been experiencing in the last decade. This research was aimed to develop a prediction model using machine learning, to tackle the problem of employee turnover in the Indian information technology industry. Although classification tools have been used to predict employee churn in the past, more focus has been given to neural networks as an effective machine learning tool and comparing the accuracies of different classification tools. In this study, an attempt has been made to test the accuracy of a simple classification tool, SVM, to predict employee attrition in IT industry, with an aim of encouraging organisations in the country to adopt it. Using current status of employee as a response variable, a SVM model was trained and its performance evaluated. Results from the feature selection showed that gender, business travel and the number of companies worked are not relevant to predict the attrition in the current study. Gender bias and glass ceiling have been known to plague the IT industry, especially in developing countries like India (Bhattacharya & Ghosh, 2012). The present model by identifying gender as an unimportant feature that can contribute to an employee leaving an IT organisation in India could suggest that the situation has improved recently.Attrition was also not seen to be influenced by the number of companies that the employees have worked in before, which signify the extent of their experience.

This internship has been a useful learning opportunity. I can say with certainty that working at DLK helped me learn a great deal and developed my fundamental competencies. It goes without saying that the technical components of the work I've done might be enhanced given enough time. I spent 47 doing research and learning about new technology. My internship taught me the value of developing my talents, being self-motivated, and expanding my knowledge.

REFERENCE

- Bogumil M. Konopka, Felicja Lwow, Magdalena Owczarz, Łukasz Łaczmański, "Exploratory Data Analysis of a Clinical Study Group: Development of a Procedure for Exploring Multidimensional Data," PLOS ONE, [Online]
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6107146/pdf/pone.0201950.pdf, August23, 2018, pp. 1-21.
- Introduction to Machine Learning using Python [Online], Available: https://www.geeksforgeeks.org/introduction-machine-learning-using-python/
- Exploratory data analysis – FromWikipedia
- https://en.wikipedia.org/wiki/Exploratory_data_analysis
- Al-Radaideh, Q.A., Al Nagi, E.: Using data mining techniques to build a classification model for predicting employees performance. Int. J. Adv. Comput. Sci. Appl. 3, 144-151 (2012)
- Chang, H.Y: Employee turnover: a novel prediction solution with effective feature selection. WSEAS Trans. Inf. Sci. Appl. 6, 417-426 (2009)

- **APPENDICES**

APPENDICES 1

INTERNSHIP BREAKDOWN

- Introduction
- Data Science and Machine Learning Concepts
- Python for Data Science • Statistics for Data Science
- Probability and Hypothesis Testing
- NumPy Data Analysis
- Panda Data Analysis
- Python Data Visualization
- Machine Learning
- Data Loading and Exploration
- Data Cleaning
- Feature Selecting and Engineering
- Linear and Logistic Regression
- K Nearest Neighbours
- Decision Trees
- Ensemble Learning and Random Forests
- Support Vector Machines
- K- means
- PCA
- Data Science Career