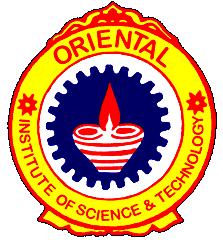
ORIENTAL INSTITUTE OF SCIENCE & TECHNOLOGY

BHOPAL

COMPUTER SCIENCE & BUSINESS SYSTEM



INTERNSHIP / TRAINING REPORT

on

Data Science and Business Analytics intern

Submitted By

Roll No: 0105CB201014

Name: Devendra Pawar

Semester: 5th

ORIENTAL INSTITUTE OF SCIENCE & TECHNOLOGY

BHOPAL (M.P.) - 462022, INDIA

Jul-Dec 2022

CERTIFICATE OF INTERNSHIP

ACKNOWLEDGEMENT

ABSTRACT

*CONTENTS*

|  |  |  |
| --- | --- | --- |
| *S.No* | *Title* | *Page No.* |
| *1* | *Learning Objective of Internship………………………………………….* | *1* |
| *2* | *Weekly overview of internship activities………………………………….* | *2* |
| *3* | *Introduction* | 3 |
| *4* | *Details of Internship* | 4 |
| *5* | *Output of Internship………………………………….……………………* | 5 |
| *6* | *Conclusion………………………………….………………………………* | 6 |
| *7* | *Bibliography………………………………….……………………………* | 7 |
|  |  |  |

LEARNING OBJECTIVES/INTERNSHIP OBJECTIVES

WEEKLY OVERVIEW OF INTERNSHIP ACTIVITIES

|  |  |  |  |
| --- | --- | --- | --- |
| 1st Week | DATE | DAY | Name of the Topic/Module Completed |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

INTRODUCTION

**Data Science-**

Data science combines math and statistics, specialized programming, advanced analytics, artificial intelligence (AI), and machine learning with specific subject matter expertise to uncover actionable insights hidden in an organization’s data. These insights can be used to guide decision making and strategic planning.

The accelerating volume of data sources, and subsequently data, has made data science is one of the fastest growing field across every industry. As a result, it is no surprise that the role of the data scientist was dubbed the “sexiest job of the 21st century” by Harvard Business Review (link resides outside of IBM). Organizations are increasingly reliant on them to interpret data and provide actionable recommendations to improve business outcomes.

The data science lifecycle involves various roles, tools, and processes, which enables analysts to glean actionable insights. Typically, a data science project undergoes the following stages:

Data ingestion: The lifecycle begins with the data collection--both raw structured and unstructured data from all relevant sources using a variety of methods. These methods can include manual entry, web scraping, and real-time streaming data from systems and devices. Data sources can include structured data, such as customer data, along with unstructured data like log files, video, audio, pictures, the Internet of Things (IoT), social media, and more.

Data storage and data processing: Since data can have different formats and structures, companies need to consider different storage systems based on the type of data that needs to be captured. Data management teams help to set standards around data storage and structure, which facilitate workflows around analytics, machine learning and deep learning models. This stage includes cleaning data, deduplicating, transforming and combining the data using ETL (extract, transform, load) jobs or other data integration technologies. This data preparation is essential for promoting data quality before loading into a data warehouse, data lake, or other repository.

Data analysis: Here, data scientists conduct an exploratory data analysis to examine biases, patterns, ranges, and distributions of values within the data. This data analytics exploration drives hypothesis generation for a/b testing. It also allows analysts to determine the data’s relevance for use within modeling efforts for predictive analytics, machine learning, and/or deep learning. Depending on a model’s accuracy, organizations can become reliant on these insights for business decision making, allowing them to drive more scalability.

Communicate: Finally, insights are presented as reports and other data visualizations that make the insights—and their impact on business—easier for business analysts and other decision-makers to understand. A data science programming language such as R or Python includes components for generating visualizations; alternately, data scientists can use dedicated visualization tools.

Data scientists rely on popular programming languages to conduct exploratory data analysis and statistical regression. These open source tools support pre-built statistical modeling, machine learning, and graphics capabilities. These languages include the following (read more at "Python vs. R: What's the Difference?"):

R Studio: An open source programming language and environment for developing statistical computing and graphics.

Python: It is a dynamic and flexible programming language. The Python includes numerous libraries, such as NumPy, Pandas, Matplotlib, for analyzing data quickly.

To facilitate sharing code and other information, data scientists may use GitHub and Jupyter notebooks.

Some data scientists may prefer a user interface, and two common enterprise tools for statistical analysis include:

SAS: A comprehensive tool suite, including visualizations and interactive dashboards, for analyzing, reporting, data mining, and predictive modeling.

IBM SPSS: Offers advanced statistical analysis, a large library of machine learning algorithms, text analysis, open source extensibility, integration with big data, and seamless deployment into applications.

Data scientists also gain proficiency in using big data processing platforms, such as Apache Spark, the open source framework Apache Hadoop, and NoSQL databases. They are also skilled with a wide range of data visualization tools, including simple graphics tools included with business presentation and spreadsheet applications (like Microsoft Excel), built-for-purpose commercial visualization tools like Tableau and IBM Cognos, and open source tools like D3.js (a JavaScript library for creating interactive data visualizations) and RAW Graphs. For building machine learning models, data scientists frequently turn to several frameworks like PyTorch, TensorFlow, MXNet, and Spark MLib.

Given the steep learning curve in data science, many companies are seeking to accelerate their return on investment for AI projects; they often struggle to hire the talent needed to realize data science project’s full potential. To address this gap, they are turning to multipersona data science and machine learning (DSML) platforms, giving rise to the role of “citizen data scientist.”

Multipersona DSML platforms use automation, self-service portals, and low-code/no-code user interfaces so that people with little or no background in digital technology or expert data science can create business value using data science and machine learning. These platforms also support expert data scientists by also offering a more technical interface. Using a multipersona DSML platform encourages collaboration across the enterprise.

Enterprises can unlock numerous benefits from data science. Common use cases include process optimization through intelligent automation and enhanced targeting and personalization to improve the customer experience (CX). However, more specific examples include:

Here are a few representative use cases for data science and artificial intelligence:

An international bank delivers faster loan services with a mobile app using machine learning-powered credit risk models and a hybrid cloud computing architecture that is both powerful and secure.

An electronics firm is developing ultra-powerful 3D-printed sensors to guide tomorrow’s driverless vehicles. The solution relies on data science and analytics tools to enhance its real-time object detection capabilities.

A robotic process automation (RPA) solution provider developed a cognitive business process mining solution that reduces incident handling times between 15% and 95% for its client companies. The solution is trained to understand the content and sentiment of customer emails, directing service teams to prioritize those that are most relevant and urgent.

A digital media technology company created an audience analytics platform that enables its clients to see what’s engaging TV audiences as they’re offered a growing range of digital channels. The solution employs deep analytics and machine learning to gather real-time insights into viewer behavior.

An urban police department created statistical incident analysis tools to help officers understand when and where to deploy resources in order to prevent crime. The data-driven solution creates reports and dashboards to augment situational awareness for field officers.

Shanghai Changjiang Science and Technology Development used IBM® Watson® technology to build an AI-based medical assessment platform that can analyze existing medical records to categorize patients based on their risk of experiencing a stroke and that can predict the success rate of different treatment plans.

Business analytics-

Business analytics, a data management solution and business intelligence subset, refers to the use of methodologies such as data mining, predictive analytics, and statistical analysis in order to analyze and transform data into useful information, identify and anticipate trends and outcomes, and ultimately make smarter, data-driven business decisions.

The main components of a typical business analytics dashboard include:

Data Aggregation: prior to analysis, data must first be gathered, organized, and filtered, either through volunteered data or transactional records

Data Mining: data mining for business analytics sorts through large datasets using databases, statistics, and machine learning to identify trends and establish relationships

Association and Sequence Identification: the identification of predictable actions that are performed in association with other actions or sequentially

Text Mining: explores and organizes large, unstructured text datasets for the purpose of qualitative and quantitative analysis

Forecasting: analyzes historical data from a specific period in order to make informed estimates that are predictive in determining future events or behaviors

Predictive Analytics: predictive business analytics uses a variety of statistical techniques to create predictive models, which extract information from datasets, identify patterns, and provide a predictive score for an array of organizational outcomes

Optimization: once trends have been identified and predictions have been made, businesses can engage simulation techniques to test out best-case scenarios

Data Visualization: provides visual representations such as charts and graphs for easy and quick data analysis

**Using business analytics tools**

Business data analytics has many individual components that work together to provide insights. While business analytics tools handle the elements of crunching data and creating insights through reports and visualization, the process actually starts with the infrastructure for bringing that data in. A standard workflow for the business analytics process is as follows:

Data collection: Wherever data comes from, be it IoT devices, apps, spreadsheets, or social media, all of that data needs to get pooled and centralized for access. Using a cloud database makes the collection process significantly easier.

Data mining: Once data arrives and is stored (usually in a data lake), it must be sorted and processed. Machine learning algorithms can accelerate this by recognizing patterns and repeatable actions, such as establishing metadata for data from specific sources, allowing data scientists to focus more on deriving insights rather than manual logistical tasks.

Descriptive analytics: What is happening and why is it happening? Descriptive data analytics answers these questions to build a greater understanding of the story behind the data.

Predictive analytics: With enough data—and enough processing of descriptive analytics —business analytics tools can start to build predictive models based on trends and historical context. These models can thus be used to inform future decisions regarding business and organizational choices.

Visualization and reporting: Visualization and reporting tools can help break down the numbers and models so that the human eye can easily grasp what is being presented. Not only does this make presentations easier, these types of tools can help anyone from experienced data scientists to business users quickly uncover new insights.

**Business analytics vs. business intelligence**

On the face of it, there may not seem to be much difference between business analytics and business intelligence. Some overlap does exist between the two, but looking at business analytics versus business intelligence still creates a gap that needs some explanation.

Certainly, the terms are extremely connected, but business intelligence uses historical and current data to understand what happened in the past and what is happening now. Business analytics, on the other hand, builds on the foundation of business intelligence and attempts to make educated predictions about what might happen in the future. In order to make data-driven predictions about the likelihood of future outcomes, business analytics uses next-generation technology, such as machine learning, data visualization, and natural language query.

**Benefits of business analytics**

Business analytics benefits impact every corner of your organization. When data across departments consolidates into a single source, it syncs up everyone in the end-to-end process. This ensures there are no gaps in data or communication, thus unlocking benefits such as:

Data-driven decisions: With business analytics, hard decisions become smarter—and by smart, that means that they are backed up by data. Quantifying root causes and clearly identifying trends creates a smarter way to look at the future of an organization, whether it be HR budgets, marketing campaigns, manufacturing and supply chain needs, or sales outreach programs.

Easy visualization: Business analytics software can take unwieldy amounts of data and turn it into simple-yet-effective visualizations. This accomplishes two things. First, it makes insights much more accessible for business users with just a few clicks. Second, by putting data in a visual format, new ideas can be uncovered simply by viewing the data in a different format.

Modeling the what-if scenario: Predictive analytics creates models for users to look for trends and patterns that will affect future outcomes. This previously was the domain of experienced data scientists, but with business analytics software powered by machine learning, these models can be generated within the platform. That gives business users the ability to quickly tweak the model by creating what-if scenarios with slightly different variables without any need to create sophisticated algorithms.

Go augmented: All of the points above consider the ways that business data analytics expedite user-driven insights. But when business analytics software is powered by machine learning and artificial intelligence, the power of augmented analytics is unlocked. Augmented analytics uses the ability to self-learn, adapt, and process bulk quantities of data to automate processes and generate insights without human bias.

Business analytics use cases

More and more departments are trying to better understand how their decisions and budgets affect the business at large. With business analytics software, it’s possible to use data to drive strategic decisions, regardless of task or department:

Marketing: Analytics to identify success and impact

Which customers are more likely to respond to an email campaign? What was the last campaign’s ROI? More and more marketing departments are trying to better understand how their programs affect the business at large. With AI and machine learning powering analysis, it’s possible to use data to drive strategic marketing decisions. Learn more

Human Resources: Analytics to find and share talent insights

What actually drives employee decisions regarding their career? More and more HR leaders are trying to better understand how their programs affect the business at large. With the right analytical capabilities, HR leaders are able to quantify and predict outcomes, understand recruitment channels, and review employee decisions en masse. Learn more

Sales: Analytics to optimize your sales

What is the critical moment that converts a lead to a sale? In-depth analytics can break down the sales cycle, taking in all of the different variables that lead to a purchase. Price, availability, geography, season, and other factors can be the turning point on the customer journey—and analytics offer the tool to decipher that key moment. Learn more

Finance: Analytics to power predictive organizational budgets

How can you increase your profit margins? Finance works with every department, be it HR or sales. That means that innovation is always key, especially as finance departments face larger volumes of data. With analytics, it’s possible to bring finance into the future for predictive modeling, detailed analysis, and insights from machine learning.

Details of Internship

TASKS-

1-Exploratory Data Analysis on retail

2- Prediction using Supervised Learning