

NAME: KADIRI, PATIENCE OTSELEMI

QUESTION 1

What is the difference between exploratory data analysis and predictive data analysis?

Exploratory Data Analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily it is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task.

While,

Predictive analytics is the branch of the advanced analytics which is used to make predictions about unknown future events. Predictive analytics uses many techniques from data mining, statistics, modeling, machine learning, and artificial intelligence to analyze current data to make predictions about future.

QUESTION 2

How would you define the role of a Data Scientist in Product Development Team?

A data scientist helps basically in the analytics. The measurements taken by metrics and the insights provided by analytics enable product teams to make informed decisions about upgrading product functionality or adding capabilities.

The roles are:

Product viability A variety of analytics tools can verify product concepts, helping developers test, learn, adjust and retest to speed up the product design and launch process.

Informed product decision-making: Analytics has made decision-making more objective, reliable, and faster. While intuition based on experience and expertise can still play a valuable role in product development, it can – and should –take a backseat to objective analytics.

Product progress measurement: Product analytics can inform team members about which features are working and which are not. Analytics plays a critical role in creating an accurate product roadmap that can tell you where your product is currently, where you want it to go (what you want it to be), and how to get it there.

User experience insights: Product teams can use analytics to understand why users are buying their product and how they are using it.

Product development inspiration: Analytics can jump-start innovation and help an existing product remain viable for an extended period of time. Quantitative analytics, used in conjunction with qualitative techniques, can provide a more holistic view of a product to help product management teams make the kind of focused improvements and adjustments that will help maintain that product's value and improve its longevity.

QUESTION 3

Outline the various phases of a typical data science methodology.

1. **Business understanding:** Every project, regardless of its size, starts with business understanding, which lays the foundation for successful resolution of the business problem.
2. **Analytic approach:** After clearly stating a business problem, the data scientist can define the analytic approach to solving it
3. **Data requirements:** Choice of analytic approach determines the data requirements.
4. **Data collection:** The data scientist identifies and gathers data resources
5. **Data understanding:** Descriptive statistics and visualization techniques can help a data scientist understand data content, assess data quality and discover initial insights into the data
6. **Data preparation:** The data preparation stage comprises all activities used to construct the data set that will be used in the modeling stage.
7. **Modeling:** Starting with the first version of the prepared data set, data scientists use a training set—historical data in which the outcome of interest is known—to develop predictive or descriptive models using the analytic approach already described. The modeling process is highly iterative.
8. **Evaluation:** The data scientist evaluates the model's quality and checks whether it addresses the business problem fully and appropriately.
9. **Deployment:** After a satisfactory model has been developed that has been approved by the business sponsors, it is deployed into the production environment or a comparable test environment.
10. **Feedback:** By collecting results from the implemented model, the organization gets feedback on the model's performance and observes how it affects its deployment environment.

QUESTION 4

Mention 4 tools that a data scientist can rely on to effectively deliver his/her work.

1. R
2. Python
3. Tableau
4. Tensor-flow