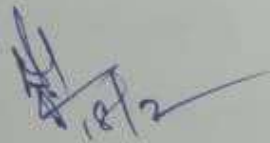


Assignment

[Data Analysis]

 18/2

by

Q. Pravin

22-CS-257

III - BSc(CS) - B

Data Analytics life cycle

→ In today's data-driven world, Data Analytics plays a vital role in helping Business to make better decisions.

→ It contains six phases each. Phase ensures that raw data is transformed into meaningful insights for effective decision making.

→ The below is an explanation of all phases with a real world example.

Real world Example: Predicting Customer churn for a Subscription Service.

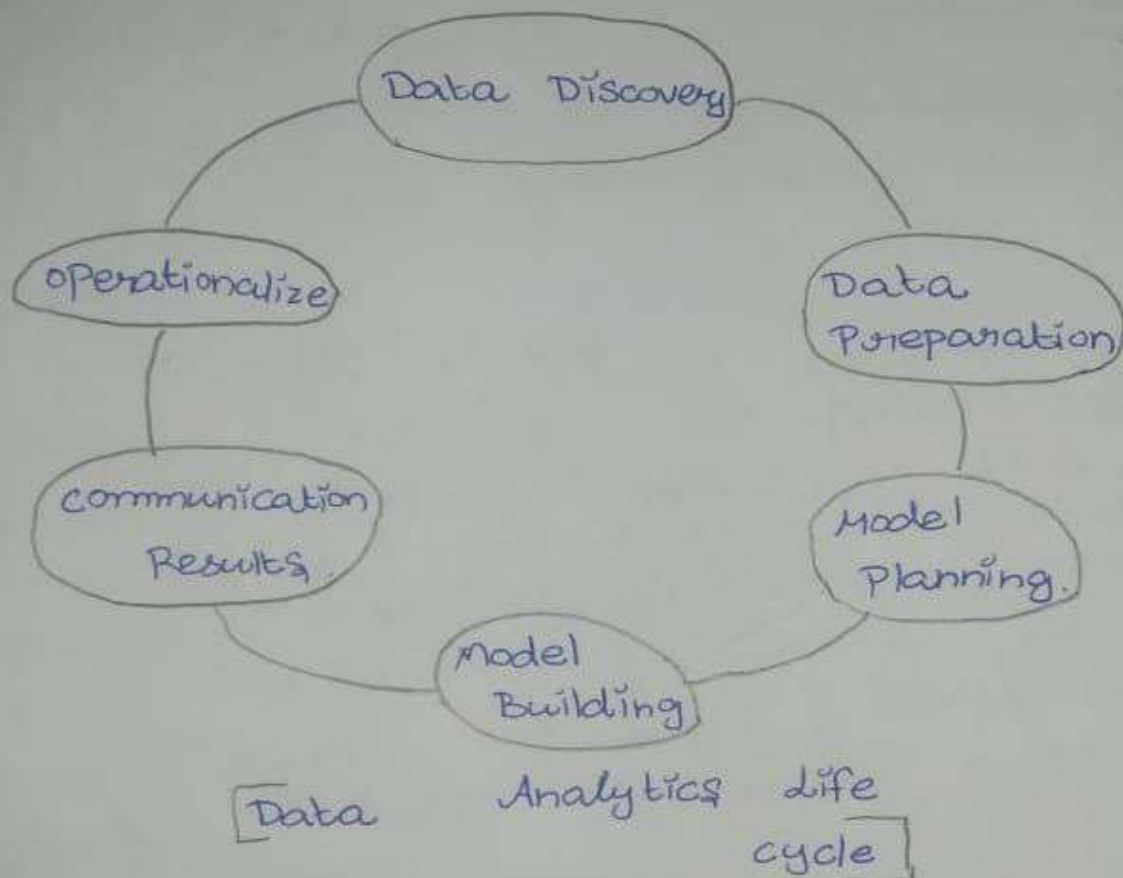
1. Data Discovery

Goal: Understand the available data and identify potential predictors of churn.

Activities

Inventory existing data sources: demographics, age, usage, Customer patterns, billing history, Customer Service interactions, website activity). Explore the

data to understand its characteristics,
 identify Potential Predictions of churn
 and define the target variable
 (what constitutes churn). Look for
 initial correlations and Patterns.



2 Data Preparation

Goal → clean, transform, and prepare the data of Modeling

Activities → clean the data by handling missing values (imputation or removal), correcting inconsistencies and removing duplications. Engineer new features.

that might be predictive (eg, average
watch time per week, number
of support tickets) Transform
data into suitable format for
modeling (scaling numerical features)

3. Model Planning

Goal → Select the appropriate modeling
techniques and plan the model building.
Process

Activities → Select an appropriate
modeling technique (eg, logistic regression,
random forest) based on the
prediction task. Define evaluation
metrics (accuracy, precision, recall,
AUC) to measure model performance
tools and technologies (Python with
scikit-learn, R, cloud platforms).

4. Model Building

Goal → Train and evaluate
the chosen model(s).

Activities → Train the chosen model
on the training data. Tune the
model's hyperparameters using the
validation set and select optimize.
Evaluate the model on the
performance.

validation Set and select the best-performing model. Finally, test the chosen model on the held-out test set to estimate its performance on unseen data.

5. communication Results

Goal → Present the findings and insights to stakeholders in a clear and understandable way

Activities → Create clear and concise visualizations (charts, graphs) to illustrate the model's performance, Predictions of churn, and key insights. Prepare a comprehensive report summarizing the analysis, including the problem definition, data description, methodology, results, and recommendations. Present the findings to stakeholders, explaining the implications for business decisions and retention strategies.

6 operationalize

Goal → Deploy the model into a production environment so it can be used to make.

the Prediction on new data.

Activities

→ Integrate the trained model into the streaming service's systems to make predictions on new customer data. Automate the model retraining process to maintain accuracy over time as new data becomes available.

→ Continuously monitor the model's performance in production and make adjustments as needed. Develop and implement targeted retention campaigns for high-risk customers based on the model's prediction.

Real World Example

"Churn for a Subscription

Service"

"Subscription - based - Streaming Service wants to Predict with which customers likely to cancel"