

Data Science Role:

Skills:

1. Strong understanding of statistical and machine learning concepts, such as probability, statistics, linear algebra, and optimization.
2. Familiarity with programming languages such as Python, R, and SQL, as well as experience using popular data science libraries and frameworks such as scikit-learn, TensorFlow, Keras, and Pandas.
3. Experience with data cleaning, preprocessing, and visualization techniques, as well as the ability to work with large and complex datasets.
4. Strong problem-solving skills, including the ability to identify and implement solutions to data-related problems.
5. Strong verbal and written communication skills, as the intern will likely be required to present their findings to both technical and non-technical audiences.
6. Familiarity with cloud-based data science platforms such as AWS, GCP, or Azure is a plus.
7. Knowledge of time series analysis, and natural language processing is a plus.
8. Familiarity with version control systems like Git is a plus.

Assignment: Predictive Modeling for Customer Churn

Objective:

The objective of this assignment is to build a predictive model that can predict customer churn for a given company. The intern will use machine learning techniques to build the model and document the process, including feature selection, model evaluation, and performance metrics.

Instructions:

1. Obtain a dataset of customer information, including demographic information, customer behavior, and whether or not the customer has churned.
2. Perform data cleaning and preprocessing on the dataset, including handling missing data and converting categorical variables to numerical variables.
3. Explore the data and perform feature selection to select the relevant features that will be used in the model.
4. Build a predictive model using machine learning algorithms such as Logistic Regression, Random Forest, or Gradient Boosting.
5. Train the model using a portion of the data and use the remaining data to evaluate the performance of the model.
6. Evaluate the model performance using metrics such as accuracy, precision, recall, F1-score and AUC-ROC.

7. Fine-tune the model by trying different parameters or techniques to improve performance.
8. Create a report detailing the process and results, including the feature selection process, the model used, the evaluation metrics, and the performance of the final model.
9. Provide a brief on the limitation of the model and the possible future work.

Deliverable: A report including the process, results, and the code used.

Note: Make sure to use appropriate techniques for handling imbalanced datasets if the data is imbalanced, and also make sure to follow best practices for avoiding overfitting, such as using techniques like cross-validation.

DataSet:

<https://www.kaggle.com/competitions/bank-marketing-uci/overview>