Data Science Methodologies

`These are notes from the coursera course Data Science Methodologies`

Week 1: From Problem to Approach and From Requirements to Collection

Module 1: From Problem to Approach

- Define the problem, metrics for success, project design trade-offs, and risks
- Determine the type of data science problem: supervised vs unsupervised learning, recommendation system, etc.
- Outline the project lifecycle stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment

Module 2: From Requirements to Collection

- Gather requirements from stakeholders
- Determine what data is needed and identify possible sources
- Evaluate feasibility of obtaining the required data
- Create a data collection plan including collection methods and storage

Week 2: From Understanding to Preparation and From Modeling to Evaluation

Module 1: From Understanding to Preparation

- Perform initial data exploration using summary statistics and visualizations
- Assess data quality and tidy the data as needed
- Identify outliers and determine how to handle them
- Determine whether any data transformations are needed

Module 2: From Modeling to Evaluation

- Based on the problem, select appropriate data modeling techniques (regression, clustering, etc)
- Train models using training data, tuning hyperparameters as needed
- Assess model performance on test data using appropriate metrics

- Analyze results to determine whether additional data or model refinements are needed
- Compare multiple models to select the best performing one

Week 3: From Deployment to Feedback

Module 1: From Deployment to Feedback

- Plan how to operationalize the model and integrate it into business processes
- Monitor the model in production to detect drift or degraded performance
- Collect user feedback to identify areas for improvement
- Re-train models periodically on new data to keep them relevant

Let me know if you would like me to expand on any part of this summary further. I'm happy to go into more detail on any of the modules or key points.