DNA Topics

The DNA Topics–with their associated subtopics–are listed below:

- Binary classification & Regression Models
    - Logistic regression
    - Random Forest
    - Support Vector Machines
    - Naive Bayes
- Neural networks
    - Backpropagation
- NLP
    - Count Vectorizer
    - TF-IDF
- Significance testing
    - T-tests
    - Mann-Whitney U
    - Standard Deviation
    - Variance
    - Z-scoring
- Unsupervised learning
    - PCA
    - t-SNE
- Model performance metrics
    - Recall
    - F1
    - Accuracy
    - Precision
- Handling imbalanced data
    - Random Oversampling
    - Random Undersampling
    - SMOTE

For each of the DNA Topics, you need to be able to answer the following questions:

1. When is this algorithm used?
2. If the DNA Topic involves algorithms, can I write out each step of the algorithm using pen and paper?
3. If an interviewer was sharing my screen, could I show them I know how to use Python to implement the DNA Topic?

If your answer is "No" to any of the above questions, then you don't know the DNA Topic. Also, you may notice that I left out possible subtopics for each DNA Topic. That was on purpose. We are only focusing on the topics that interviewers will ask during your data scientist interview.