**Syllabus for cycle 1:**

This is your syllabus for 1st week along with some resources and a **mini task** will be released each day to test your understanding and hands-on skills.

**Day 1: Data Cleaning & EDA**

**Syllabus:**

* Missing values, outliers
* EDA using pandas & plots

**Resources:**

* [Data Cleaning with Pandas](https://towardsdatascience.com/data-cleaning-with-python-using-pandas-library-c6f4a68ea8eb/)
* [Handling missing values](https://www.kaggle.com/code/alexisbcook/missing-values)
* [Outliers](https://www.geeksforgeeks.org/data-science/detect-and-remove-the-outliers-using-python/)
* [Data Visualization](https://www.kaggle.com/learn/data-visualization)

**Day 2: Feature Engineering + Regression**

**Syllabus:**

* Encoding (one-hot, ordinal), scaling, normalization
* Feature selection (correlation, mutual info)
* Linear, Ridge, Lasso Regression
* Evaluation metrics - MSE, R² metrics

**Resources:**

* [Scaling & Normalization](https://www.kaggle.com/code/alexisbcook/scaling-and-normalization)
* [Handling Categorial variables](https://www.kaggle.com/code/alexisbcook/categorical-variables)
* [Mutual information for feature selection](https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_classif.html)
* [Linear regression](https://www.geeksforgeeks.org/machine-learning/linear-regression-python-implementation/)
* [Regularization methods](https://towardsdatascience.com/understanding-l1-and-l2-regularization-93918a5ac8d0/)
* [Lasso and Ridge](https://www.kaggle.com/code/ishaanthareja007/basic-linear-ridge-and-lasso-regression)
* [See only Mse, Mae and R square](https://scikit-learn.org/stable/modules/model_evaluation.html#regression-metrics)

**Day 3: Classification + Evaluation Metrics**

**Syllabus:**

* Logistic Regression
* Linear discriminant analysis
* Classification metrics: Accuracy, ROC-AUC

**Resources:**

* [Logistic regression youtube vedio](https://www.youtube.com/watch?v=yIYKR4sgzI8)
* [Logistic regression tutorial](https://www.kaggle.com/code/mnassrib/titanic-logistic-regression-with-python)
* [LDA](https://www.youtube.com/watch?v=azXCzI57Yfc)
* [LDA vs PCA](https://www.kaggle.com/code/saptarsi/lda-using-python)
* [Accuracy, presicion, recall](https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall)
* [AUC and ROC](https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc)

**Day 4: Tree-Based Models + k-NN + Feature Selection**

**Syllabus:**

* Decision Trees (structure, criteria)
* Random Forests (feature importance, hyperparams)
* k-NN
* Drop weak features using importance/Mutual info

**Resources:**

* [Decision Trees](https://www.geeksforgeeks.org/machine-learning/decision-tree-introduction-example/)
* [Decision Trees-Youtube vedio](https://www.youtube.com/watch?v=7VeUPuFGJHk)
* [Feature importance with random forests](https://www.geeksforgeeks.org/dsa/random-forest-classifier-using-scikit-learn/)
* [Random forests](https://www.geeksforgeeks.org/machine-learning/random-forest-algorithm-in-machine-learning/)
* [Knn- youtube vedio](https://www.youtube.com/watch?v=HVXime0nQeI)
* [knn](https://www.geeksforgeeks.org/machine-learning/k-nearest-neighbor-algorithm-in-python/)

**Day 5: Model Comparison + Summary Table**

**Syllabus:**

* Accuracy, ROC-AUC comparison
* Overfitting risk
* Finalizing a model pipeline

**Resources:**

* [Overfitting, underfitting](https://www.kaggle.com/code/dansbecker/underfitting-and-overfitting)
* [Classifier comparison](https://www.kaggle.com/code/sinakhorami/titanic-best-working-classifier)

**Day 6: Final Challenge – Dirty Dataset Detective**