



DAILY WORK
REPORT
TR-02

INFOWIZ

19 JUNE 2024

Day 13: Ensemble Methods - Random Forests

Summary: Today, we delved into ensemble methods, specifically focusing on Random Forests, a powerful ensemble learning technique widely used for classification and regression tasks. We explored how Random Forests improve upon decision trees, their implementation using Python with scikit-learn, and their applications in real-world scenarios.

Key Learnings:

1. Introduction to Ensemble Methods:

- Defined ensemble methods as techniques that combine predictions from multiple individual models to improve overall performance and robustness.
- Discussed advantages such as reducing overfitting, capturing complex relationships in data, and enhancing predictive accuracy compared to individual models.

2. Random Forests:

- **Composition:** Explained that Random Forests are comprised of multiple decision trees, each trained on a subset of the dataset (bootstrap samples) and using random subsets of features (feature bagging).
- **Voting Mechanism:** Described how Random Forests aggregate predictions through averaging (regression) or voting (classification) of individual tree predictions.

3. Implementation in Python:

- Preprocessed the dataset to handle missing values, encode categorical variables, and ensure data readiness for Random Forest modeling.
- Utilized scikit-learn's RandomForestClassifier and RandomForestRegressor classes to build Random Forest models for classification and regression tasks, respectively.
- Tuned hyperparameters such as number of trees (n_estimators), maximum depth of trees, and minimum samples per leaf to optimize model performance.

4. Practical Application:

- Applied Random Forests to a real-world dataset, such as predicting customer preferences or diagnosing medical conditions based on patient data.
- Explored feature importance rankings generated by Random Forests to identify key predictors influencing model predictions.

Today's session provided a comprehensive understanding of Random Forests as a robust ensemble learning technique, capable of handling complex datasets and improving predictive accuracy through model aggregation. The practical implementation in Python equipped us with valuable skills to apply Random Forests in diverse machine learning applications.