

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Факультет інформатики та обчислювальної техніки
Кафедра інформаційних систем та технологій

Лабораторна робота №4

з дисципліни «Програмування інтелектуальних інформаційних систем»

Виконав:

студент групи ПІ-11

Лисенко Андрій

Київ, 2023

1. Тюнінг Гіперпараметрів (На основі [Hyperparameter tuning tutorial](#))

Мета

Покращення якості моделей машинного навчання шляхом оптимізації гіперпараметрів.

Використані моделі та результати

```
LogisticRegression - Best Parameters: {'C': 0.1, 'solver': 'liblinear'},  
Accuracy: 0.9912280701754386  
GaussianNB - Best Parameters: {}, Accuracy: 0.9649122807017544  
DecisionTree - Best Parameters: {'max_depth': 5, 'min_samples_leaf': 2,  
'min_samples_split': 10}, Accuracy: 0.9473684210526315  
RandomForest - Best Parameters: {'max_depth': 7, 'min_samples_leaf': 1,  
'min_samples_split': 2, 'n_estimators': 200}, Accuracy: 0.9649122807017544  
SVM - Best Parameters: {'C': 0.1, 'gamma': 'scale', 'kernel': 'linear'},  
Accuracy: 0.9824561403508771  
KNN - Best Parameters: {'algorithm': 'auto', 'n_neighbors': 7, 'weights':  
'distance'}, Accuracy: 0.9473684210526315
```

Comparison with Default Parameters:

```
LogisticRegression - Default Parameters Accuracy: 0.9736842105263158  
GaussianNB - Default Parameters Accuracy: 0.9649122807017544  
DecisionTree - Default Parameters Accuracy: 0.9473684210526315  
RandomForest - Default Parameters Accuracy: 0.9649122807017544  
SVM - Default Parameters Accuracy: 0.9736842105263158  
KNN - Default Parameters Accuracy: 0.9473684210526315
```

- **Logistic Regression:**

- Найкращі параметри: {'C': 0.1, 'solver': 'liblinear'}
- Точність: 99.12%
- Точність з дефолтними параметрами: 97.36%

- **GaussianNB:**

- Точність: 96.49% (немає параметрів для тюнінгу)
- **Decision Tree:**
 - Найкращі параметри: {'max_depth': 5, 'min_samples_leaf': 2, 'min_samples_split': 10}
 - Точність: 94.73%
- **Random Forest:**
 - Найкращі параметри: {'max_depth': 7, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200}
 - Точність: 96.49%
- **SVM:**
 - Найкращі параметри: {'C': 0.1, 'gamma': 'scale', 'kernel': 'linear'}
 - Точність: 98.24%
 - Точність з дефолтними параметрами: 97.36%
- **KNN:**
 - Найкращі параметри: {'algorithm': 'auto', 'n_neighbors': 7, 'weights': 'distance'}
 - Точність: 94.73%

Model	Tuned Accuracy	Default Accuracy
LogisticRegression	0.9912	0.9737
GaussianNB	0.9649	0.9649
DecisionTree	0.9474	0.9474
RandomForest	0.9649	0.9649
SVM	0.9825	0.9737
KNN	0.9474	0.9474

Висновки

Тюнінг гіперпараметрів дозволив підвищити точність моделей, особливо для Logistic Regression та SVM. Це підтверджує ефективність алгоритмічного підбору гіперпараметрів.

2. Ансамблі та Бустінг (На основі Ensemble learning techniques)

Мета

Дослідження різних методів ансамблювання моделей для підвищення загальної точності прогнозування.

Використані методи та результати

Technique	Accuracy
Voting	0.964912
Averaging	0.964912
Weighted Averaging	0.973684
Stacking	0.973684
Blending	0.982456
Bagging	0.956140
Boosting	0.973684
Tuned Decision Tree	0.929825

Висновки

- **Ансамблі** (Voting, Averaging, Weighted Averaging) використовують різні моделі для досягнення більш стабільних та точних прогнозів.
- **Стекінг** та **Блендінг** ефективно комбінують прогнози кількох моделей, використовуючи мета-модель для остаточного прогнозу.
- **Бустінг** поступово зменшує помилки за допомогою послідовного навчання моделей.
- **Беггінг** зменшує варіативність та підвищує стабільні