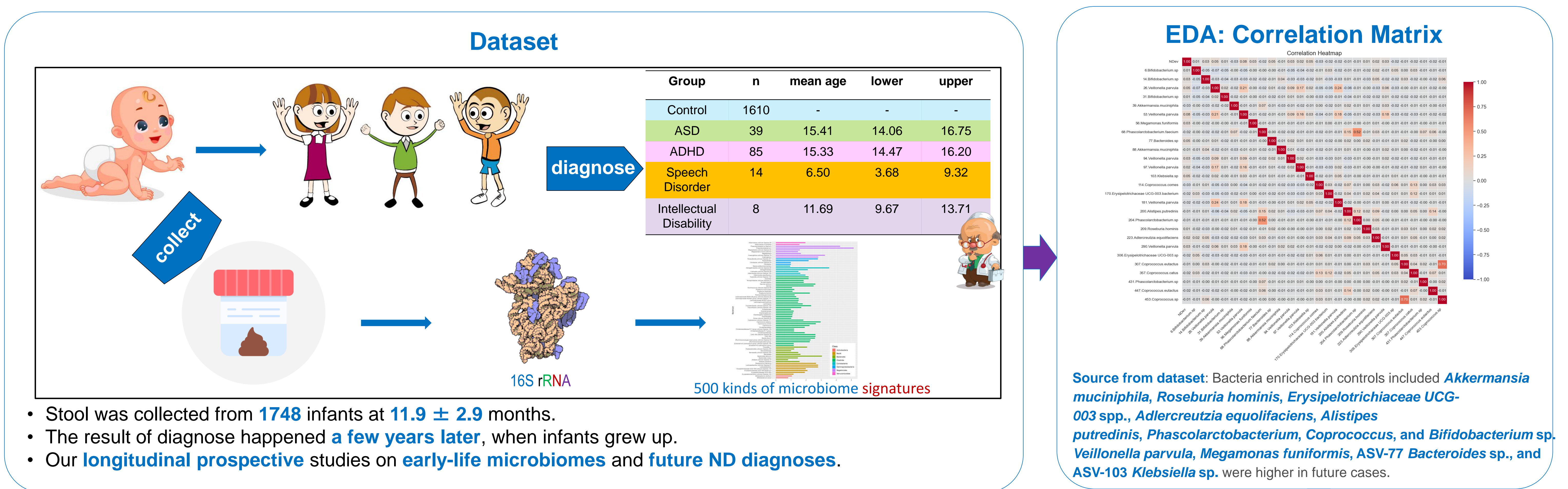


**CENTENNIAL**  
**COLLEGE**

<sup>1</sup> Artificial Intelligence - Software engineer technology, Centennial College

STEM  
Fellowship

- Neurodevelopmental disorders (NDDs) affect brain and nervous system development.
- Alterations in infant gut microbiome biomarkers would be associated with their future NDDs.
- However, prospective early life data of infant gut microbial biomarkers are limited.
- Based on prospective data, I develop predictive models of infant future NDDs with present gut microbial biomarkers, which can foresee their future NDDs outcomes.



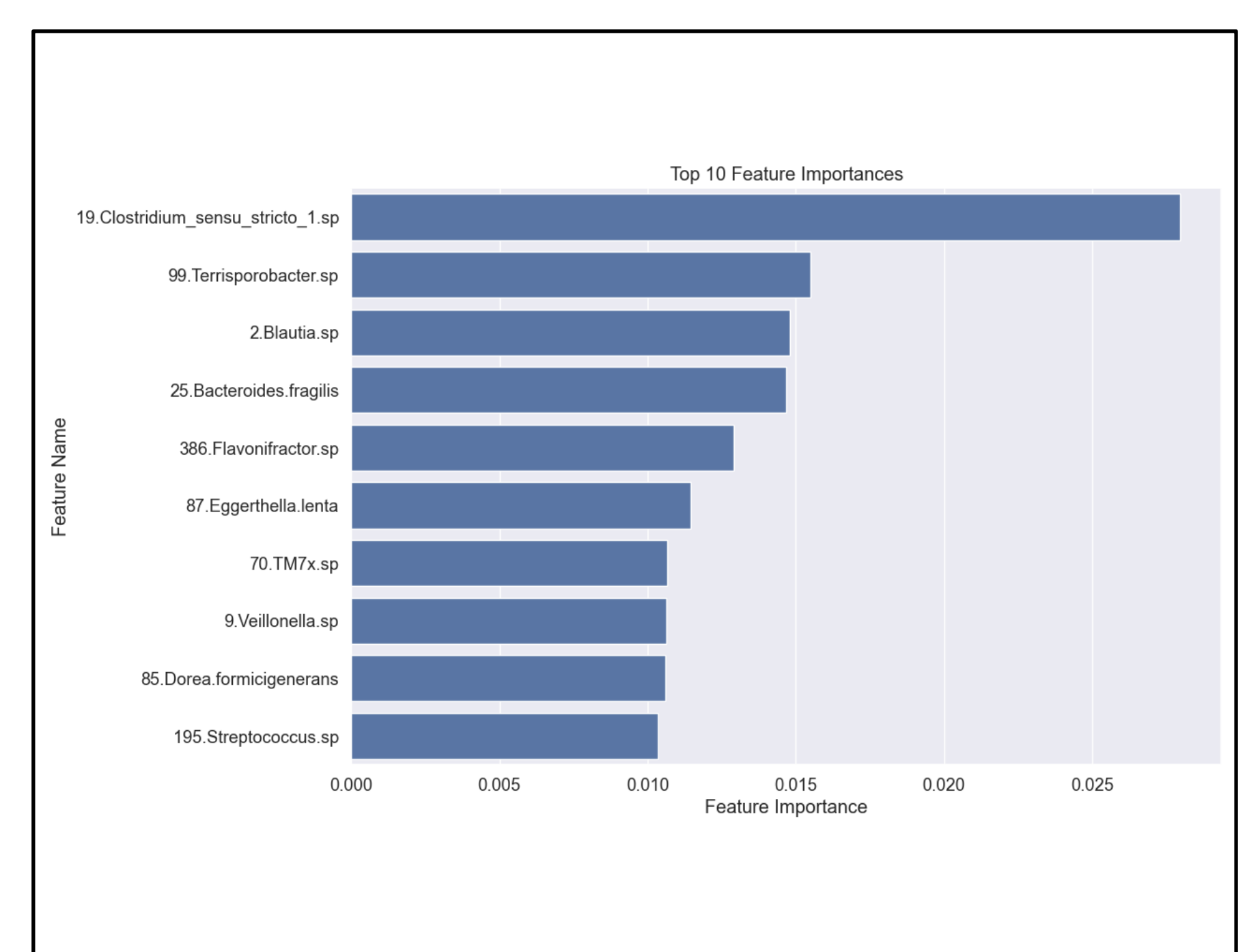
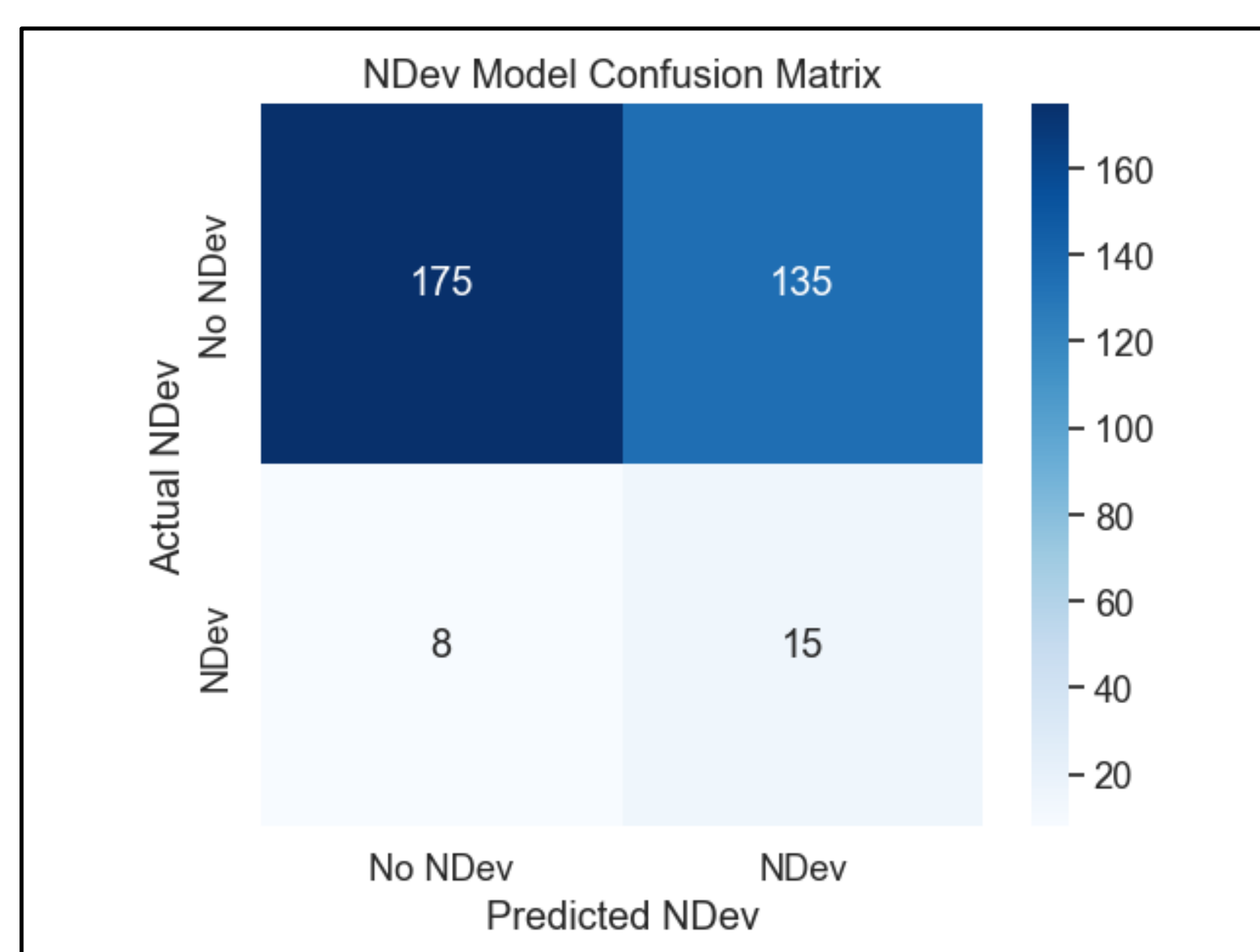
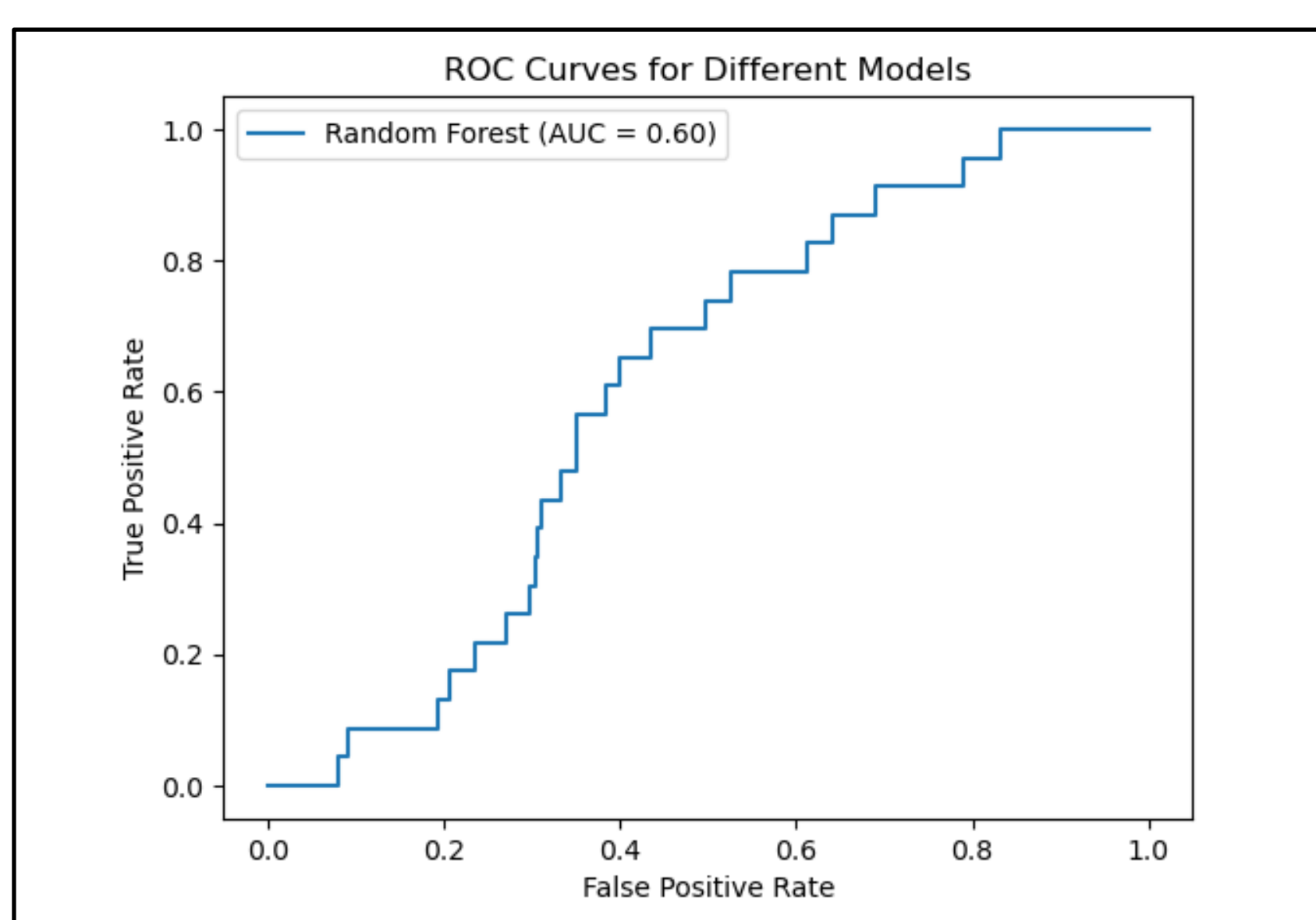
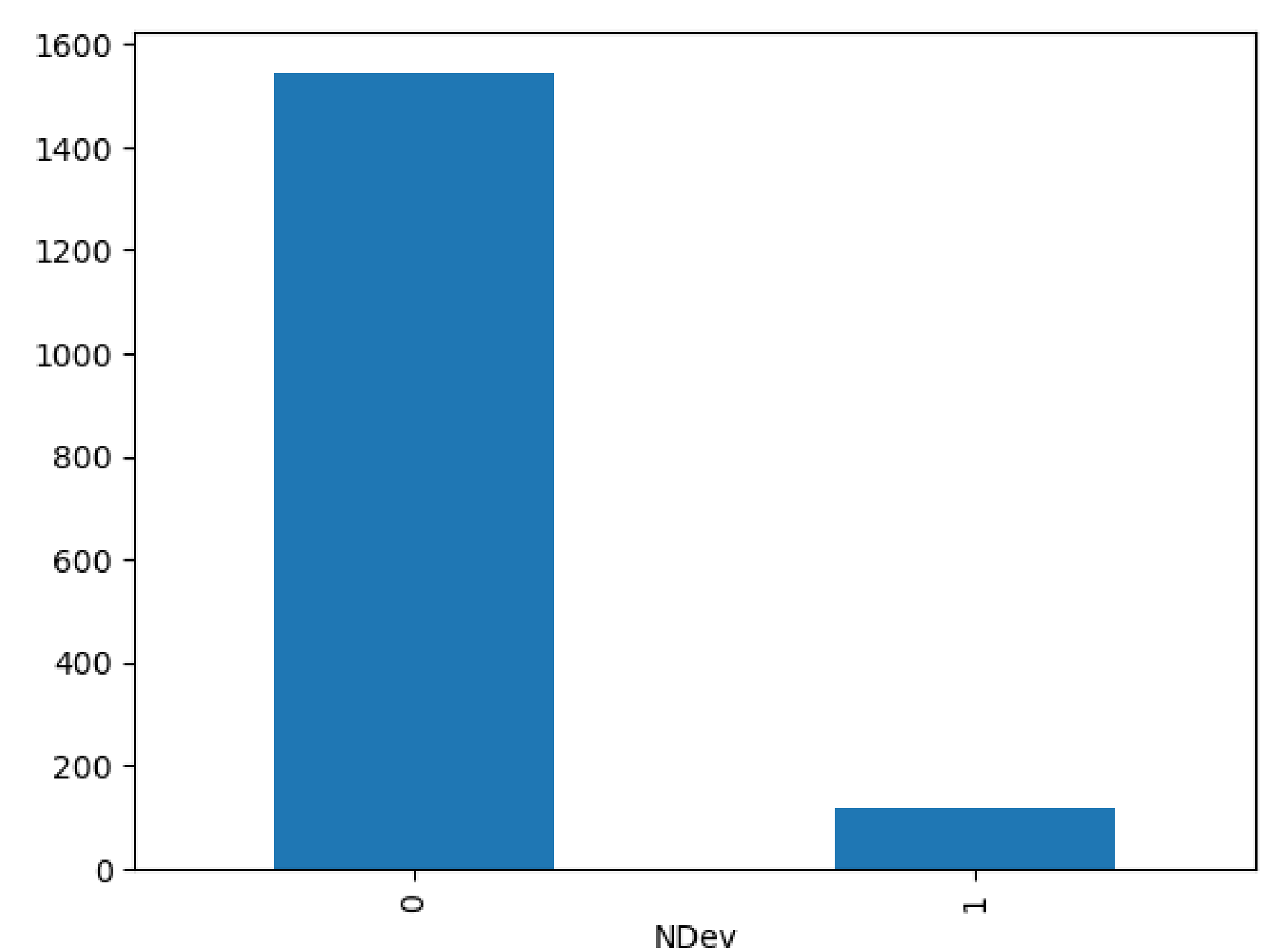
The screenshot shows a Jupyter Notebook interface. On the left, a sidebar lists files and folders: 'EnsembleLearning.ipynb', 'GradientBoostingClassifier', 'GradientBoostingRegressor', 'HistGradientBoostingClassifier', 'HistGradientBoostingRegressor', 'IsotonicForest', 'RandomForestClassifier' (highlighted), 'RandomForestRegressor', 'RandomizedSearchCV', 'StackingClassifier', 'StackingRegressor', 'VotingClassifier', and 'VotingRegressor'. The main area displays the code for the `RandomForestClassifier` class from the `sklearn.ensemble` module. The code includes a docstring and several parameters: `n_estimators=100`, `criterion='gini'`, `max_depth=None`, `min_samples_split=2`, `min_samples_leaf=1`, `min_weight_fraction=0.0`, `max_features='log2'`, `max_leaf_nodes=None`, `min_impurity_decrease=0.0`, `bootstrap=True`, `oob_score=False`, `oob_score_loss='squared_error'`, `random_state=None`, `verbose=0`, `warm_start=False`, `class_weight=None`, `ccp_alpha=0.0`, `max_samples=None`, `montecarlo_criterion=None`, and `verbose=0`. A comment indicates that `verbose=0` is the default. Below the code, a paragraph explains that the random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. It also mentions that the sub-sample size is controlled with the `max_samples` parameter and that `bootstrap=True` (default) controls whether the dataset is used to build each tree. A final note states that for a comparison between the `RandomForestClassifier` and `HistGradientBoostingClassifier`, one should refer to the `User Guide`.

Original dataset

Sample of **majority** class

**Original Positive Class** : 7%

**Undersampling Positive Class** : 49.2%



- Limitations of this study include few samples of future NDDs.
- Further studies focus on improving the model prediction and predicting specific future NDDs subtypes( ASD, Intellectual Disability, Speech Disorder, or ADHD) in the larger cases datasets.