Just-In-Time Software Defect Prediction

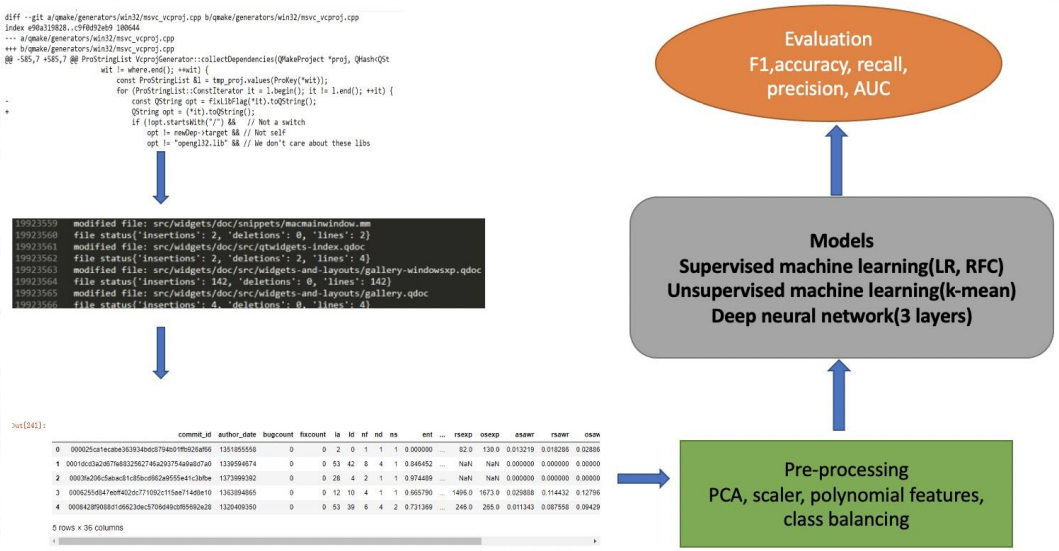
Ali Mohamed, Yifei Gong, Yunhua Zhao

Project Category: binary **classification** for **buggy** commits

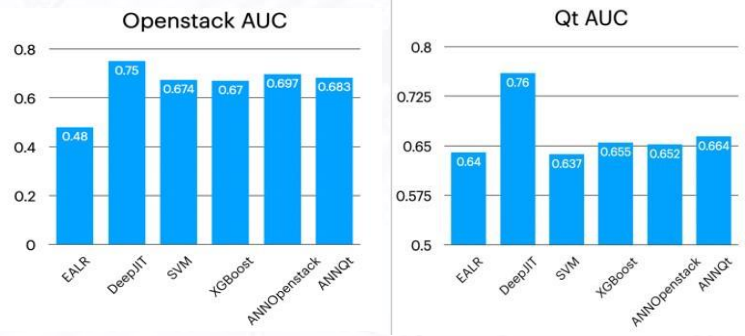
Data: two datasets:

- **Openstack** and **Qt**
- Openstack (10658 negative samples, 1616 positive samples), Qt (23148 negative samples, 2002 positive samples).
- Each sample contains information from a GitHub commit.
- Each sample has 35 features extracted from that commit (lines modified, developer experience, etc.) and 1 target values (buggy or not).

Goals: compete with state-of-the-art work by using handcrafted features



Hypothesis 1: two datasets may share intrinsic similarities that can render better results when trained together

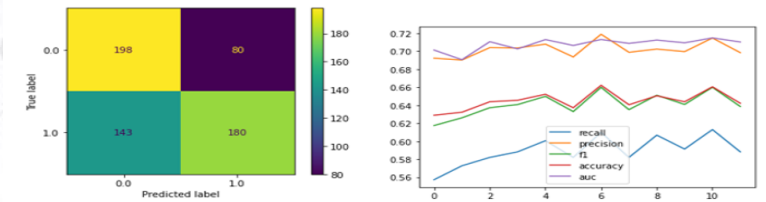


EALR and DeepJIT are state-of-the-art work . The rest four models are trained on combined oversampled datasets. ANNOpenstack and ANNQt uses feed forward neural network.

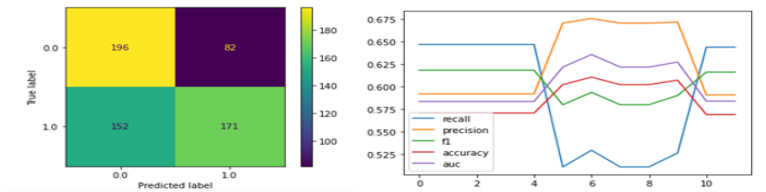
- 4 models can outperform EALR
- Still a gap with DeepJIT

Comparing the 2 main Supervised Machine Learning Models

Random Forest:



Logistic Regression:



The Random Forest Model seems to be slightly better, partially due to the size of the dataset – performance difference can be attributed to size.