



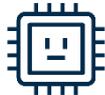
This poster introduces applied machine learning techniques for quality assurance of regulatory data, specifically applied to credit unions. The objective is to apply machine learning techniques to classify data points into two categories, raise or don't raise for plausibility review. A random forest machine learning algorithm is selected for the classification problem.

THE PROBLEM

THE VOLUME OF DATA REQUIRED FOR MONETARY, FINANCIAL STABILITY AND MICRO-PRUDENTIAL POLICY PURPOSES CONTINUES TO GROW



OUR SYSTEMS AND TOOLS FOR CHECKING DATA ARE BECOMING OBSOLETE, AS IS THE MODEL USED FOR CHECKING THE DATA



WHAT WE'VE DONE

BUILT A MACHINE LEARNING ALGORITHM FOR PREDICTING ERRONEOUS DATA POINTS

INTEGRATED A DECISION CAPTURER TO RECORD AND STORE ALL DECISIONS TO ENHANCE MODEL PERFORMANCE

QUARTERLY MODEL CALIBRATION TO IMPROVE MODEL PREDICTIVE PERFORMANCE



Before ...



Each reporting round 40,000 data points are processed, of which approximately 4,000 are classified as 'raise' for analyst to review as part of the review process

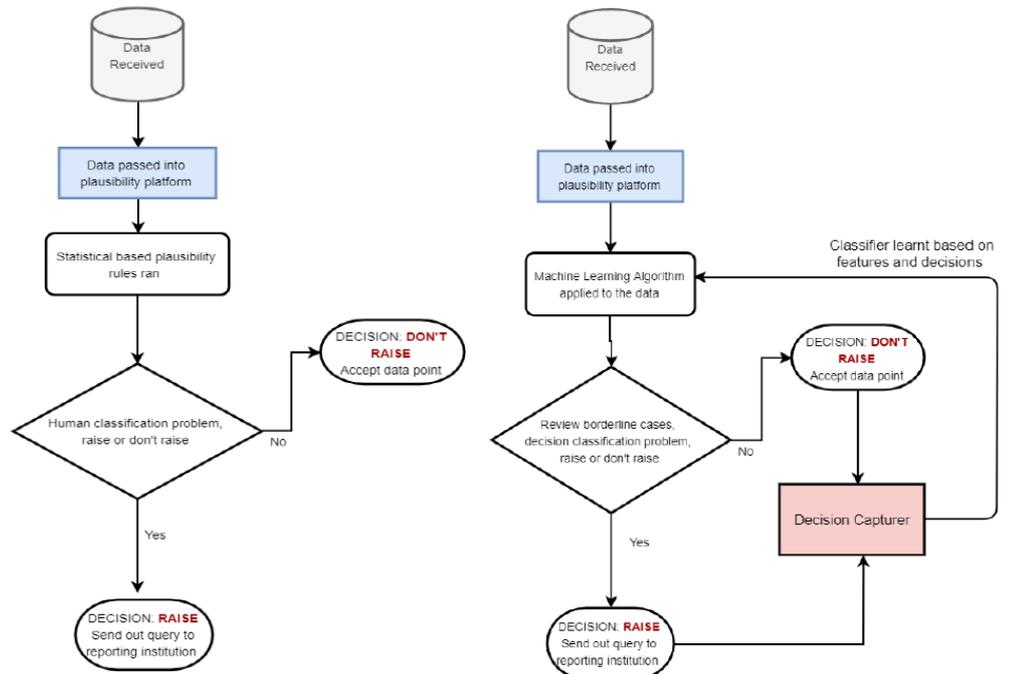


Previous techniques to classify data points were based on statistical methods such as measures of centrality and dispersion. Of the data points raised using statistical techniques, approximately 10% are classified 'correctly'



The ability to precisely classify observations is valuable for the regulatory data checking process as it reduces the time spent manually reviewing 'correct' data items. The choice of classifier is the random forest classifier

Old process → New process



MODEL PERFORMANCE

The model is predicting with 92% accuracy, with a balanced accuracy of 75%

Using the decision capturer to record decisions, model performance is improving with each calibration round

| Confusion Matrix | Truth | | |
|------------------|-------|--------|----|
| | No | Yes | |
| Predicted | No | 19,040 | 68 |
| | Yes | 771 | 81 |