# Charge Classification Engine Release Notes

# Date – 5th December, 2016

# Objective – Algorithm tightening, bounded learning time and database updates

1. Added a second syntactic classifier based on Damerau-Levenshtein distance algorithm that complements the Jaro algorithm. Now the classification engine has a total of 4 algorithms, 2 syntactic (Jaro and Damerau-Levenshtein) and 2 semantic (Logistic Regression and Neural Network), described as follows:
   1. **Jaro Distance** – Jaro distance is a measure of similarity between two strings. The idea is that if you calculate the number of transpositions between each of the two given strings and the number of different characters, then that results in a distance or a number between 0 and 1. The closer the number is to 1, the more similar the two strings are, 1 being a perfect match and 0 being no match.
   2. **Damerau-Levenshtein Distance –** The Levenshtein distance is a measure of the amount of differences (edits) between two strings. Differences include minimal number of insertions, deletions and replacements needed for transforming one string into another. Additionally in Damerau-Levenshtein distance, transposition of adjacent symbols is also allowed.
   3. **Multinomial Logistic Regression –** Multinomial Logistic Regression is a linear regression analysis to where the dependent variable is nominal with more than two levels. Thus it is a polytomous (multi-level) logistic regression.
   4. **Neural Network –** Neural Network is a non-linear modeling approach that consists of a number of simple, highly interconnected processing elements. It process information using dynamic state responses to external inputs and outputs a classification.
2. Tightened the algorithms by:
   1. Boosting the confidence scores for Jaro, Damerau-Levenshtein & Logistic Regression algorithms.
   2. Changing the ensemble voting methodology to prioritize 3-algorithms-agree at the same level as all-algorithms-agree.
3. Bounded learning time by windowing the historical training data.
4. Added a new table called CC\_ALGORITHM\_RESULTS that writes output charge classification and confidence score for each claim line along with the individual charge classification and confidence score of each of the 4 algorithms.