

MACHINE LEARNING for PRE-AUCTION SAMPLE SELECTION

Indika Hiran Wijesinghe
Informatics Institute of Technology
Sri Lanka
hiran@pureceylontea.com

Dr. Sameera D Viswakula
Department of Statistics
University of Colombo
sam@stat.cmb.ac.lk

Abstract- A Scientific method of selecting all possible fraudulent combinations from the weekly pre-auction, where tea samples are bided by the tea exporting companies to be purchased for exporting purposes is critical to mitigate frauds in tea exports from Sri Lanka. Therefore, building a proper mechanism to predict fraudulent samples received by the tea tasting unit of the Sri Lanka Tea Board is a challenge for the government's regulatory/testing body. The main focus of this paper is to identify a suitable pre auction sample selection algorithm with Machine Learning approach to develop a self-learning sampling system to pick the most likely fraudulent samples for physical testing.

I. INTRODUCTION

Tea is a one of the major businesses in Sri Lanka. It has become a country icon and a well-known export product for generations since the 1880s. Sri Lanka Tea Board is the government's regulatory body that regulates the Tea Industry of Sri Lanka. Tea Board was established in 1976, and it comprises both government and private sector bodies attached to the Tea sector [1]. Sri Lanka has the world's largest orthodox black tea auction which was established in the year 1883 [2]. Now it has become backbones of the national economy. In 2015 an income of Rs.182.1Bn which accounts for 54% of total Agricultural exports was brought by the tea [3]. Sri Lanka has exported around 25000 Metric Tons per month in both bulk and value-added format in year 2016[4]. Sri Lanka is the world's largest orthodox tea exporter and the 3rd largest tea exporter having 16.9 % of market share [5].

Sri Lanka Tea board is the statutory board for Tea industry regulating 8 brokers, 21 plantation companies, 352 exporters, 710 tea factories and 242

The performance impact of tea tasting data feature set is demonstrated on Locally Weighted Learning, Logistic Regression, Random Forest, C4.5 Naïve Bayes and Bayesian Network. Though classification accuracy between algorithms is similar, computational performance can differ ominously. Thus, we then show that it is worthwhile differentiating algorithms based on computational performance rather than on classification accuracy alone.

Keywords- *Supervised Learning, Classification; Machine Learning; Machine Learning Techniques; ML based model performance; ML based Model accuracy, Tea Industry*

warehouses. Presently, approximately 2 Million people, 10% of the population are engaged with the industry [5] directly or indirectly. So, tea industry has become a part of life of a significant percentage of the workforce in Sri Lanka, and therefore, it has a social & economic value for the country.

II. PROBLEM DOMAIN

Due to the fact that there is a huge demand for 'Ceylon Tea', certain tea exporters and tea brokers adulterate tea with various elements with the intention of grabbing a big stake of the profit. Therefore, the management has observed numerous frauds within the industry since the last decade, which has led to the tarnishing of the good name of this royal brand. In addition, according to Figure 1, in year 2006, Sri Lanka had a 19.6% of market share in the world tea export market. But, it decreased by 3% in the year 2015. Nevertheless, China, Vietnam and Kenya, the other tea exporting countries which had market shares of 18.1%, 6.6% and 19.8% respectively have increased their market shares 18.2%, 6.7% and 24.8% respectively. So, the