PROPOSAL

Jesse Wood, Bing-Xue, Mengjie Zhang, Bach Hoai Nguyen, Daniel Killeen

Victoria University
Engineering and Computer Science
Kelburn, Wellington, New Zealand

ABSTRACT

Submitted in partial fulfilment of the PhD in Artificial Intelligence.

***Index Terms*—** Feature Selection, Gas Chromatography, Support Vector Machines, Visualisation

1. INTRODUCTION

- Scope - place the problem in the world.
- Specifics to New Zealand, sustainability.
- Fish processing - automation, quality control, contaimination.
- Current state-of-the-art
  - GC-MS, manual, time consuming, expensive, destructive, instrumental drift.

2. LITERATURE

- Mass spectrometry [1]
- REIMS
- Classification
- Feature Selection
- Interpretable ML
- Genetic Programming

3. PRELIMINARY WORK

- Automated Fish Classification on GC-MS data.
- Genetic Programming (GP) for GC-MS data
  - Single-Tree GP
  - Multi-tree GP
- REIMS exploratory data analysis

Thanks to New Zealand Plant & Food Research for datasets, funding and expertise.

4. CONTRIBUTIONS

- Each research question applies to the Hoki and Jack Mackeral datasets.
- For each dataset, hoki and mackeral.

These are the research questions from Plant and Food Research.

- Can REIMS data be used to classify different fish tissues? What variables are responsible?
  - Classification
  - Feature Importance - Interpretable
- Can REIMS data detect mineral oil contamination in fish? At what concentration? What variables are responsible?
  - Classification
  - Regression
  - Feature importance - Interpretable
- Can REIMS data be used to distinguish between different fish individuals? What variables are responsible?
  - Identification
  - Feature imporance - Interpretable

5. MILESTONES

6. THESIS OUTLINE

1. Introduction
2. Background
   - Mass Spectrometry
   - REIMS
   - Classifcation / Regression
   - Interpretable ML

3. Exploratory Data Analysis
4. Preprocessing
5. Classification
6. Contaminiation Detection
7. Individual Identification
8. Auto ML
9. Discussion
10. Conclusion

7. RESOURCES

- Hardware
  - ECS Grid Compute
  - Rapoi
  - Niwa HPC - via Auckland University
- Software
  - Repository - Github
  - Project Management - Github Projects
  - Programming language - Python
  - Documentation - Read the Docs
- Experience
  - Field-trip to Callaghan Innovation to see REIMS
  - Field-trip to NZ Plant and Food Research (if necessary for future datasets).

8. APPENDIX

9. REFERENCES

[1] K Eder, "Gas chromatographic analysis of fatty acid methyl esters," *Journal of Chromatography B: Biomedical Sciences and Applications*, vol. 671, no. 1-2, pp. 113–131, 1995.