



VICTORIA UNIVERSITY OF  
**WELLINGTON**  
TE HERENGA WAKA

**School of Engineering and Computer Science**  
Te Kura Mātai Pūkaha, Pūrorohiko

PO Box 600  
Wellington  
New Zealand

Tel: +64 4 463 5341  
Internet: [office@ecs.vuw.ac.nz](mailto:office@ecs.vuw.ac.nz)

**Rapid detection and identification  
of contamination within fish  
products using rapid ionisation  
evaporative mass spectrometry**

Jesse Wood

Supervisors: Bach Hoai Nguyen, Bing Xue, Mengjie  
Zhang

Submitted in partial fulfilment of the requirements for  
Doctorate of Philosophy - Artificial Intelligence.

**Abstract**

This document gives some ideas about how to write a project proposal, and provides a template for a proposal. You should discuss your proposal with your supervisor.



## 1. Introduction

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

## 2. Literature Review

- Mass spectrometry [?]
- REIMS
- Classification
- Feature Selection
- Interpretable ML
- Genetic Programming
- Transfer Learning

## 3. Preliminary Work

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

## 4. Contributions

- Each research question applies to the REIMS and Data Infusion .
- For each dataset, hoki and mackerel.

These are the research questions from Plant and Food Research.

- Can REIMS data be used to classify different species tissues? What variables are responsible?
  - Classification
  - Feature Importance - Interpretable

- Can REIMS data detect mixed-species contamination in fish tissues? At what concentration? What variables are responsible?
  - Classification
  - Regression
  - 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 importance - Interpretable

## 5. Milestones

- Literature Review
- EDA
- Preprocessing
- Classification
- Cross-species Contamination
- Mineral-oil Contamination
- Individual Identification
- Auto ML
- Thesis

## 6. Thesis Outline

1. Introduction
2. Background
  - Mass Spectrometry
  - REIMS
  - Classification / Regression
  - Interpretable ML
3. Preparations

- Exploratory Data Analysis
- Preprocessing

#### 4. Applications

- Classification
- Contamination Detection
- Individual Identification
- Auto ML

#### 5. Discussion

#### 6. Conclusion

## 7. Resources

In this section you will detail any resource requirements such as hardware, software or access to subjects.

- 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).



# Bibliography