# CYSF Logbook

Project: Machine Learning for Predictive Maintenance

Student: Tyson Sze

Partner: Decibel Cannabis Company

Dates: September 2025 to April 2026

**Week 1: Project Setup and Data Collection**

**Dates: September 1-7, 2025**

- Contacted Decibel about project
- Discussed data availability
- Received data
- Set up python environment
- Started working with data

Notes: Data looks complete and good, some missing values, will have to do some processing.

**Week 2: Research and Planning**

**Dates: September 8-21, 2025**

- Literature review on other projects
- Studied random forest, xgboost, and lightgbm
- Made hypothesis
- Planned feature engineering approach
- Identified evaluation metrics

Notes: Making good progress, might add a fourth model if time permits.

**Week 4-5: Data Preprocessing**

**Dates: September 22 - October 5, 2025**

- Loaded data into pandas
- Edited data to remove missing columns
- Created target variables
- Created temporal train/test split
- Verified data quality

Notes: Class imbalance of 32% yes / 68% no downtime.

**Week 6-7: Feature Engineering**

**Dates: October 6-19, 2025**

- Created 5 efficiency metrics
- Created 5 temporal features
- Created 5 lag features
- Created 3 rolling averages
- Created 9 production metrics

Notes: Total of 27 features engineered.

**Week 8-9: Random Forest Models**

**Dates: October 20 - November 3, 2025**

- Implemented random forest classifier
- Tuned parameters

- Classification: 90.3% accuracy, 94% recall
- Added random forest regressor
- Regression: 7.8 MAE, r squared .71

Notes: Random forest performed very well.

**Week 10-11: XGBoost Models**

**Dates: November 4-17, 2025**

- Implemented xgboost classifier
- Edited gradient boosting parameters
- Classification: 87.1% accuracy, 88% recall
- Added xgboost regressor
- Regression: 9.1 MAE, r squared .64

Notes: XGBoost did worse than random forest - surprising.

**Week 12-13: LightGBM Models**

**Dates: November 18 - December 8, 2025**

- Implemented lightgbm classifier
- Used smote for class imbalance
- Classification: 96.8% accuracy, 100% recall
- Added lightgbmt regressor
- Regression: 10.3 MAE, r squared .59

Notes: Perfect recall is great! Hypothesis wrong since lost in regression.

**Break: Holiday Period**

**Dates: December 9, 2025 - January 5, 2026**

Notes: Took break for the holidays, reviewed results and prepared for analysis.

**Week 14-15: Feature Importance/Analysis**

**Dates: January 6-19, 2026**

- Found feature importance from all models
- Rolling average - 64% importance
- Created accuracy metrics charts
- Created feature importance charts
- Analyzed why random forest was able to win regression

Notes: Rolling averages by far the most predictive.

**Week 16-17: Economic Impact Analysis**

**Dates: January 20 - February 2, 2025**

- Consulted Decibel on current repair costs
- Calculated baseline costs
- Projected savings with models
- Validated assumptions with staff

Notes: ROI could be very strong.

**Week 18-21: Research Paper Writing**

**Dates: February 3-28, 2026**

- Outline structure
- Wrote most sections
- Documented all results and analysis
- Completed sections
- Final edits and proofreading

Notes: 14 500 word research paper complete.

**Week 22: Online Submission**

**Dates: March 1-4, 2026**

- Completed CYSF online forms
- Uploaded description, etc
- Created project images
- AFilled in all text

Notes: Also made video/slides

**Onwards: Completion**

**Dates: March 4 - April 9, 2026**

- Finish making poster board
- Finish in person presentation prep
- Prepare for Q and A
- Send full report with materials to Decibel
- Create elevator pitch