Snapshot Report

AUSTRALIAN MEAT PROCESSOR CORPORATION

Machine Learning for Retention

Machine Learning as a tool to improve retention (Stage 1)

Project Code	Prepared by	Date Submitted
2022 - 1138	Katie Scholten	21/12/2022

Project Description

In this project, Toustone sought to explore the possibility of analysing HR data to predict employee retention risk with the expectation of generating insights that will enable active intervention thus raising staff retention rates and significantly reducing costs to industry associated with this.

The focus was on the available Human Resources (HR) data from one selected red meat processor who agreed to supply data and work with us on developing a model. Further data metrics can be introduced into the model at later date to enhance and improve the model's accuracy.

By analysing, preparing data and building a Machine Learning (ML) model that learns from behaviours of past employees, with the intention to identify like patterns in current employees. The ML model will be instrumental in providing early insights to HR and team managers that allow a proactive conversation to occur and any issues to be addressed well in advance of any unplanned absenteeism or departure.

Two project objectives were agreed upon. These included:

1. To understand if Machine Learning can assist in predicting the likelihood of an employee leaving and hence be a tool to reduce turnover
2. Can Machine Learning identify venerable employees before they leave an organisations?

A further objective was added through a project variation after the success of the first two objectives. This was:

3. To apply the model to two further test sites and thus prove the model's viability and transferability between processors.

Project Content

The scope was defined as exploring retention triggers by employees at two Australian based red meat processing plants owned by a prominent red meat processor. The dataset supplied was for the period 2019 to 2021 and separated into 4 processing dates from which 'as of' analysis could be modelled and included 12 months of data prior to and 12 months of data after each processing date. Data metrics identified and provided were:

- Employee ID
- Gender
- Year of birth
- Employment type
- FTE
- Home postcode
- Leave hours taken
- Employment start date
- Employment
- Employee location
- Employee site
- Visa-Citizenship status
- English as a second language

A binary experiment question was agreed as likely to produce the strongest results given the available data.

Experiment question:

Predict the likelihood that an employee leaves within 12 months

Possible outcomes:

Employee will not be employed after 12 months

Employee will be employed after 12 months

Six algorithm models were explored;

- Logistical regression;
- Random Forest;
- KNN;
- Decision Tree classifier;
- CatBoost; and
- XGBoost.

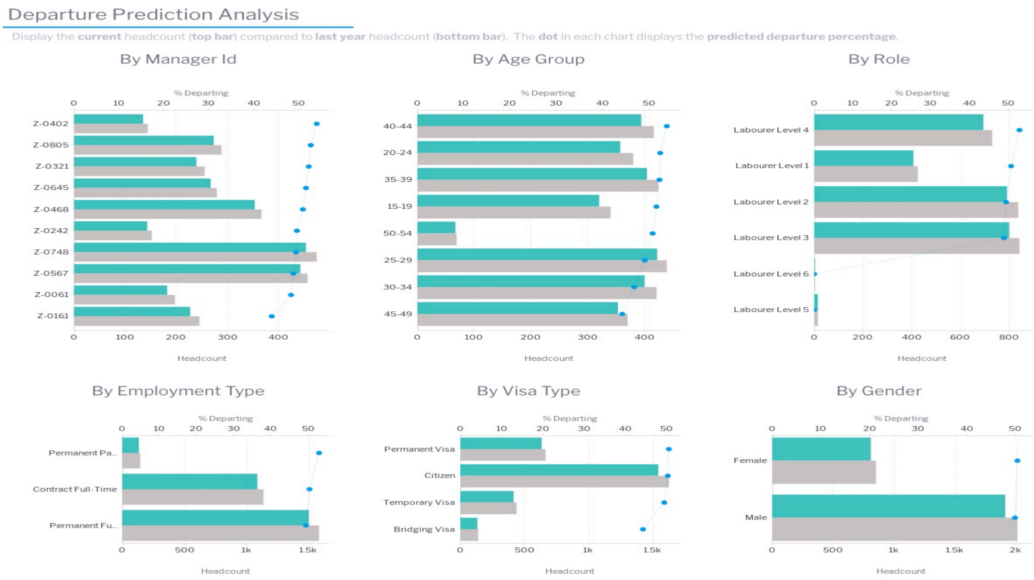
XGBoost was determined to be the most accurate. With strong results found for the Primary Processor stakeholder, a further 2 processors were invited, and datasets provided. Datasets between each processing site varied slightly and the model was adjusted accordingly.

Project Outcome

The Machine Learning models found:

	Primary Processor	Secondary Test Processor 1	Secondary Test Processor 2
% of the employees who left within 12 months that the model predicted (Recall)	59%	70%	87%
% of which the model accurately predicted those identified as at risk in the recall (Precision)	84%	80%	83%
Overall F1 score (fit of the model, 0 – poor to 1 - perfect)	0.69	0.75	0.85

As a part of best practice requirements, a reporting dashboard in Yellowfin was created to provide a visualization of the data and its relationship to the outcomes discovered in the Machine Learning model. The dashboard included reports examining current and predicted headcount, departure prediction analysis, turnover analysis, leave analysis, departures vs new arrivals analysis and comparative analysis as well as detailed reports highlighting the characteristics of identified employees at risk of leaving within 12 months of analysis. An example of the BI reports is included below.



Current At Risk Analysis

Display the current list of at risk employees based on specific filters.

1 - 100 / 144

Site	Location	Manager	Team	Team Leader	Current Role	Employee Id	ML Outcome	Employment Type	Gender	Visa Type	Age Group	Leave Hours
Western Plant	W_Skins	Z-0805	WS1	A-0761	Labourer Level 1	I-4760	At Risk	Permanent Full-Time	Male	Bridging Visa	20-24	152.00
						O-0766	At Risk	Permanent Full-Time	Female	Citizen	45-49	106.40
						P-4760	At Risk	Contract Full-Time	Male	Temporary Visa	20-24	0.00
						X-0765	At Risk	Permanent Full-Time	Male	Permanent Visa	20-24	706.80
					Labourer Level 2	A-0768	At Risk	Permanent Full-Time	Male	Citizen	45-49	311.60
						E-0765	At Risk	Permanent Full-Time	Male	Citizen	35-39	608.00
						F-0763	At Risk	Contract Full-Time	Male	Citizen	35-39	790.40
						G-2760	At Risk	Contract Full-Time	Female	Bridging Visa	35-39	45.60
						H-0761	At Risk	Permanent Full-Time	Male	Citizen	25-29	258.40
						H-0767	At Risk	Contract Full-Time	Female	Citizen	15-19	912.00
						I-0766	At Risk	Permanent Full-Time	Male	Permanent Visa	20-24	182.40
						L-4761	At Risk	Permanent Full-Time	Male	Permanent Visa	45-49	190.00
						M-0764	At Risk	Contract Full-Time	Female	Permanent Visa	20-24	440.80
						R-0760	At Risk	Contract Full-Time	Male	Permanent Visa	20-24	243.20

With the completion of the model customization and the dashboards, an opportunity to explore the Machine Learning outcomes in a way which would provide greater insight into the individuals identified as ‘at risk’ presented itself.

The initial feature analysis undertaken at the time of the model development showed which metrics contributed the most to the algorithm for each organisation. This highlighted in order of importance which metrics added the most weight to determining ‘at risk’ individuals. Using this as a base, 7-8 metrics were selected to undertake further analysis. Each metric was categorised as either continuous or categorical and analysed accordingly.

Continuous variables, tenure, age, and base rate of pay, were found to fit a curve which saw probability of leaving reduce as the variable value increased. Analysis of leave taken in the previous 12 months was also analysed for Secondary Test Processor 1. For each organisation and each continuous variable, a fitting function was produced enabling the calculation for the probability for the variable at the individual employee level, highlighting how important that variable is to the determining of an employee being identified as leaving or not leaving within the next 12 months.

Categorical variables, cost centre, employment type, employment site, and contract houses per week, were analysed in turn and a risk rating from Highest to Lowest applied to each category of each variable.

What does this mean for business?

- Identifies employees at risk of leaving (list of employee IDs produced)
- Provides an opportunity to work directly with at risk employees
- Provides insight into what increases the likelihood of employees leaving
- Provides insight into why employees choose to stay
- Enables data driven decisions on changes and initiatives in the business that can be implemented to reduce employee turn over
- Saves money – recruitment, processing, and productivity costs
- The model is transferable to different operations and meat processors with limited customization required provided the core data metrics with sufficient history is provided.

Benefit for Industry

The outcomes of the Machine Learning and dashboards were presented to each Red Meat Processor in turn. Discussions are continuing with one processor in particular keen to explore the viability of operationalizing the model. Another processor is interested in the concept but would like to see a longer-term application of the model through a 12-to-18-month implementation trial before committing to an ongoing operationalization. AMPC support for an implementation trial would be sought should the processor wish to continue to the next stage. Both organisations are keen to broaden the dataset metrics to further enhance the model outcomes.

With AMPC assistance, the next step is to establish a group of 5 plants of various size and structure to deploy the Machine Learning model to. This will not only support the processors in addressing retention issues but also enable ongoing research and refinement of the model. Any Human Resource metrics acquired will also be incorporated, with consent, into AMPC's benchmarking model for the continuing benefit of the industry.

Successful deployment of the model would not only assist the processing sector but also have carry over benefit throughout the rest of the red meat supply chain. If adopted widely within the sector, outcomes could further support producer returns and make the Australian red meat industry more competitive globally.