Cost Estimation for Manufacturing

Shane Carmichael, MBA, Bachelor of Mechatronic Engineering

Agenda

Key topics covered in this presentation

- Executive Summary
- Background and Nature of Business Problem
- Project Workflow
- Dataset, Cleaning and Feature Engineering
- Machine Learning Accuracy
- Business Predictions and Outcomes

Executive Summary

PROBLEM

How can we use machine learning to improve the speed and accuracy in cost estimation?

DATASET

Historical Cost Dataset for bespoke commercial HVAC equipment

STRATEGY

Create machine learning model from historical cost data to provide better estimates for future sales.

MODELS

Linear Regression
Lasso Regression
Ridge Regression
Bagging with LR
Deep Learning
Random Forest
ADAboost
XG Boosting

RESULTS

Predicted increase in gross profit of up to 22% while reducing overhead and working capital

BUSINESS PROBLEM

Gross Profit Margin is Inconsistent Due to Inaccurate Cost Estimates

Quantitative Cost Estimation Techniques are Slow and Expensive

Production Capacity is Limited by Available Labour

Need to Avoid Under and Overpricing

Business Question

How can we quickly and cost effectively estimate the cost of projects to avoid over and under pricing?

5

Prior Research

Has this been done before?

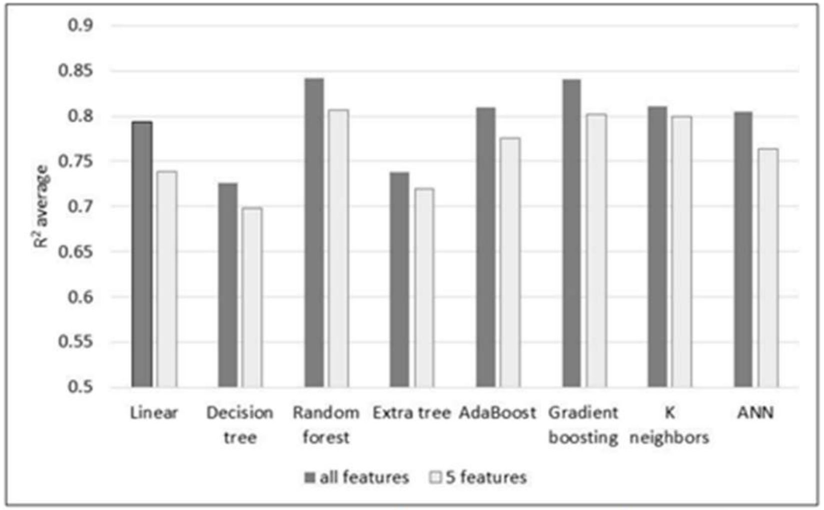


Fig. 3. Average values of determination coefficient R^2.

6

Rickenbacher et al. (2013)

Linear Regression CNC machined parts.

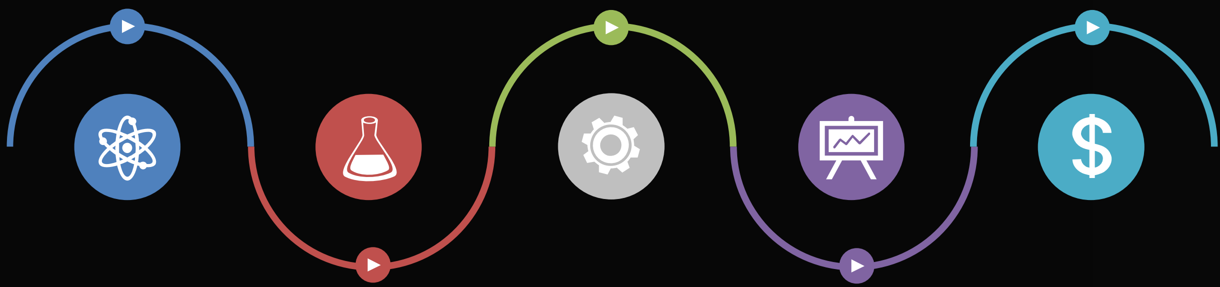
Duran et al. (2012)

ANN for cost estimation for shell and tube heat exchangers.

Kurasova et al (2021)

Multiple machine learning model in customized furniture production.

Project Workflow



Clean Data

Clean and merge datasets

Feature Engineering

Create new features using Regex

Create Models

Create ML models from dataset

Get Predictions

Use model on historical data to obtain predictions

Evaluate

Quantify the benefit of the predictions

Data Set

Data, Cleaning and Feature Engineering

```python
with pd.option_context('display.max_colwidth', 1000):
    print (dfjobs.iloc[3,4])
```

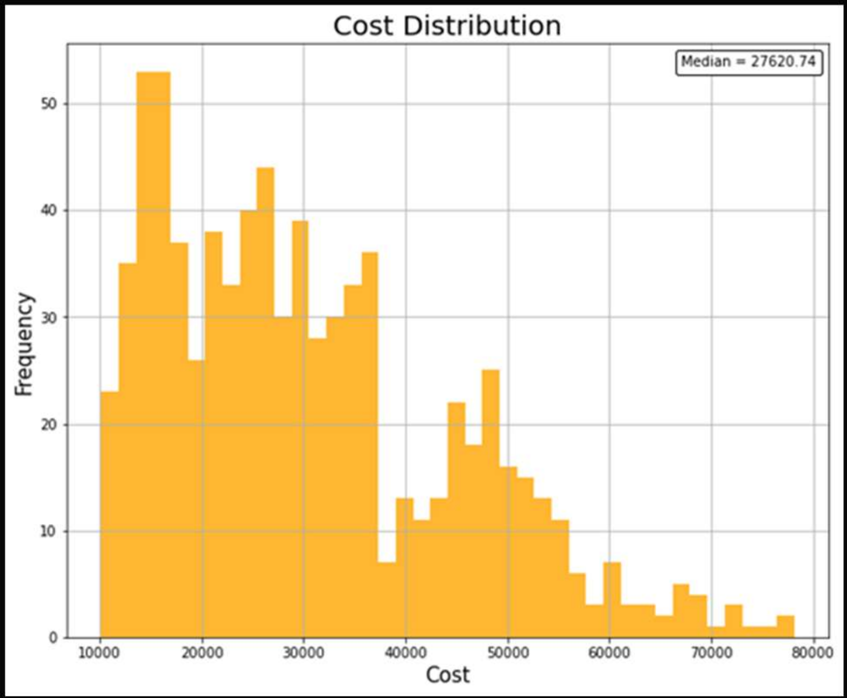
```
125 kW Rooftop Packaged Unit with Dual Stage Heating and Cooling
SA : 5000 l/s, OA : 2625 l/s

- 75% Total efficiency Enthalpy heat Exchanger
- Double-skinned 50mm PIR Sandwich Panel Cabinet
- Corrosion Treated Coils
- EC Plug Fans c/w proprietary air volume controllers
- Remote condenser
- Economy Cycle (can be operated as night purge)
- Bypass Dampers - Can be utilised for pull-down cycle or CO2 control
- Functionality wired to a terminal strip for control by others
- 12 Months Labour Warranty
- Commissioning Assistance included as standard
```

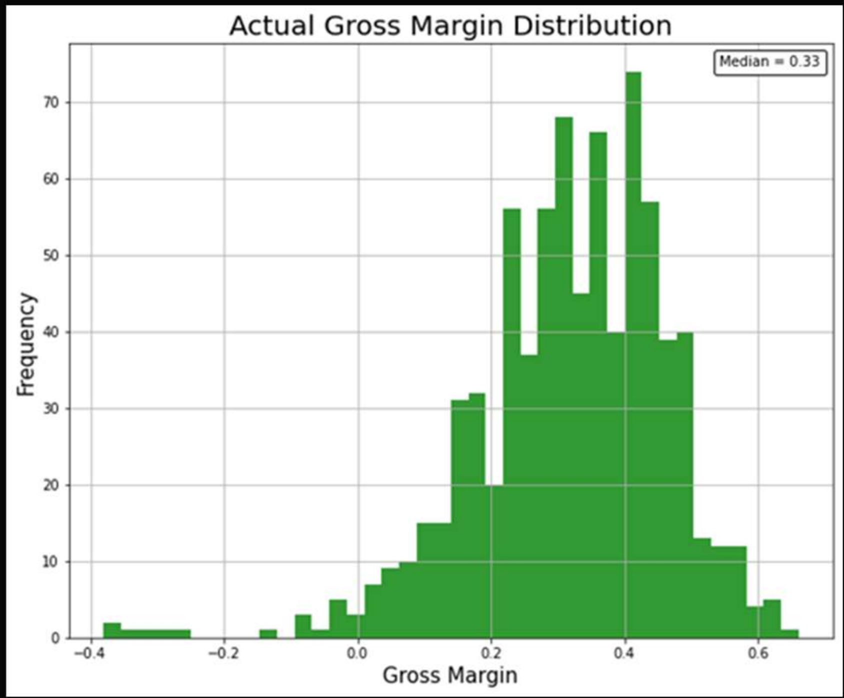
- Historical Costing Dataset
- Regex to extract key information from descriptive text
  - Capacity (kW)
  - Airflow (l/s)
  - Economy Option
- Add additional features
  - Completed Date (age)
  - Part Group
- Converted null values
- Merged costs from jobs table
- Remove outliers
- Normalise data
- Create dummy variables

EDA

Cost and GM% Distribution 5 Year



9



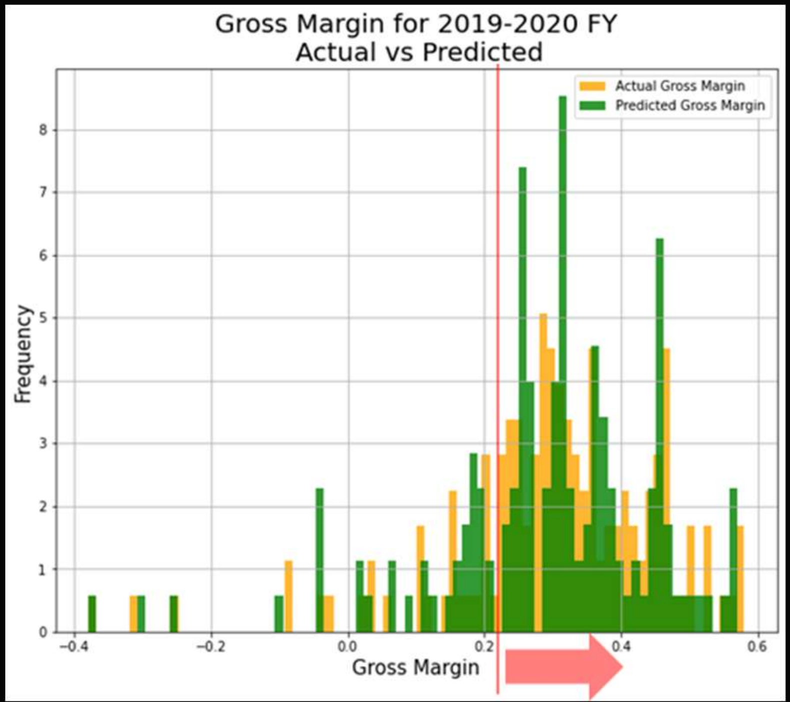
Machine Learning Results

Choose XG boost as it has the best performance

Model Name	R2 Train Score	R2 Test Score	RMSE Test	MAPE Test %
Linear Regression	0.73	0.74	7347.87	18.92
Ridge	0.73	0.74	7345.93	18.83
Lasso	0.73	0.74	7347.78	18.92
Bagging with LR	0.72	0.72	7599.72	18.71
AdaBoost	0.67	0.62	8914.31	22.69
XGBoost	1.00	0.86	5359.45	14.28
Random Forest	0.97	0.86	5445.20	14.35
Deep Learning	0.74	0.75	7195.01	53.66

Results

Applying the Model

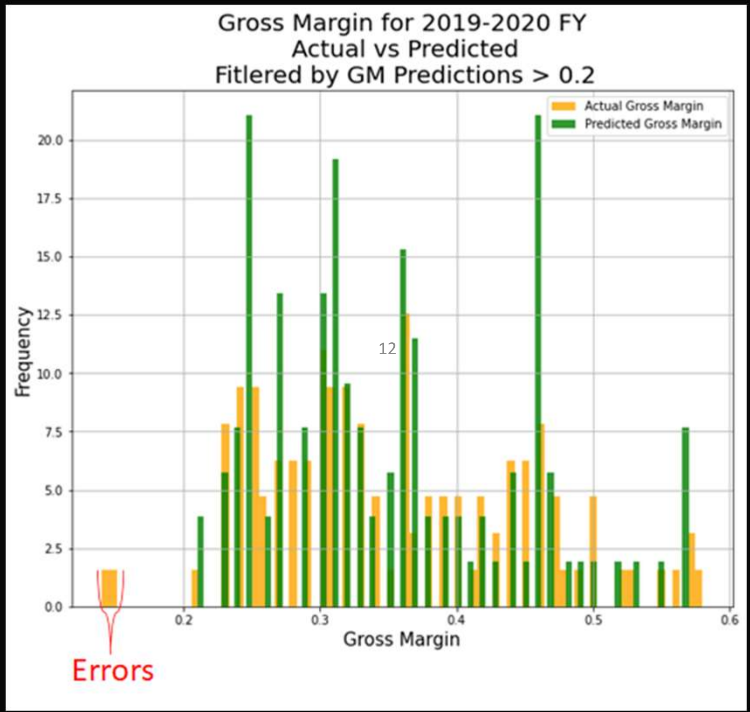


11

- Apply the model to a specific financial year (2019-2020 FY) to predict cost and gross margin
- Choose threshold for minimum GM = 0.2 (reasoning explained later)
- Select only projects with predicted GM greater than threshold to simulate how the model would work in practice
- Evaluate new GP, GM, savings in labour hours (providing additional production capacity)

Results

Gross Margin Actual Versus Predicted



Results

2019-2020 Financial Year

	Actual	After applying predictions and filtering GM > 20%	Delta Change
Revenue	$8.35M	$7.13M	- $1.22M
Costs	$5.58M	$4.52M	-$1.06M
Gross Profit	$2.77M	$2.61M	- $0.16M
Gross Margin %	33%	37%	+ 4%
Labour Hours	31,180	25,504	- 5676

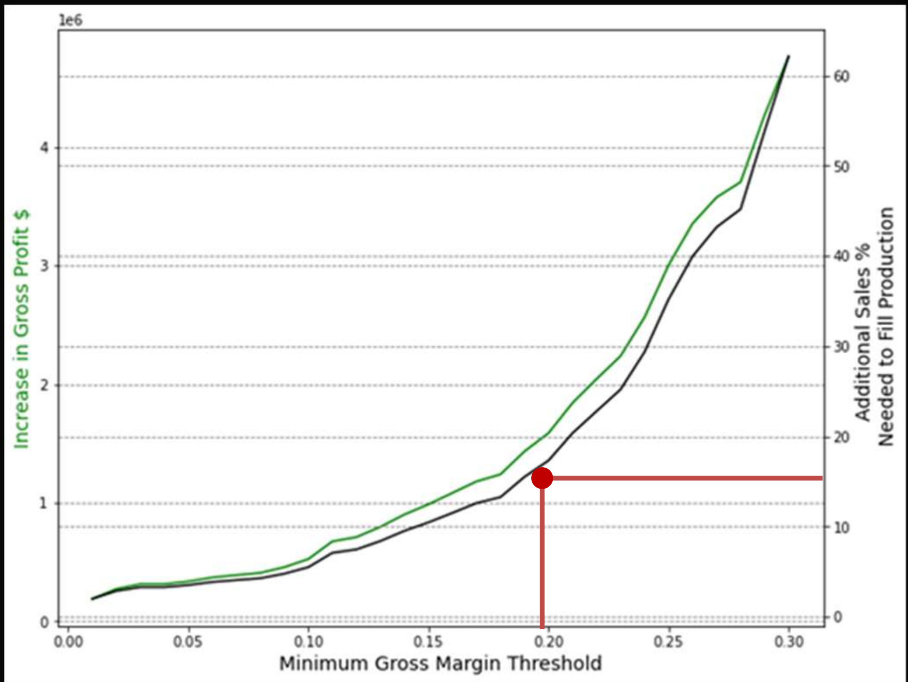
13

Financial Year	Labour Hours Saving	Additional Productive Capacity	Increased Gross Profit
2019 - 2020	5676	22.5%	$580,835
2018 - 2019	3346	16.9%	$336,624
2017 - 2018	4235	21.3%	$525,014

- Save 5676 labour hours by selecting only 'good projects'
- Represents a 22.25% increase in available production capacity
- Assuming other orders can be taken at the average gross margin, the new gross profit would be $3.18M which is a $580,835 improvement
- There are additional benefits
  - Lower working capital
  - Less overhead
- Scalable

Results

GM Threshold 2017-2021



14

- GP increases with greater GM threshold
- Need to be able to replace rejected orders with others, limited by potential sales
- Data can be created to calculate this limit by tracking lost orders specifically due to the inability to produce on time.
- Assume 15% additional sales with experience, intuition and guessing.
- Equivalent to a margin criteria of approximately 0.2

CONCLUSION AND FUTURE DEVELOPMENTS

Equipment costs can be estimated using machine learning models

Accurate costings can increase profits by eliminating poor projects

Need to investigate under/overpricing on yet to be awarded projects

Place more engineering rules in constructing the model to improve accuracy

15

Thankyou