RELATIONSHIP BETWEEN CLIMATE CHANGE AND CORPORATE FINANCIAL PERFORMANCE

The University of Melbourne
Group 22

MAST90017 Data Science Project 2

OVERVIEW

1. **Introduction**
   - Host Organisation
   - Group Members
   - Prologue
2. **Project**
   - Stages of the Data Science Pipline
   - Challenges
3. **Dataset Discussion**
   - Financial Datasets
   - Climate Datasets
   - Joining Datasets

OVERVIEW

4. **Literature/Techniques Review**
   - Related Literature Areas
   - Suitability of Approaches to Our Work
5. **Methodology**
   - Features Engineering
   - Regression Model & Hypothesis
   - Classification Model
6. **Results and Conclusion**
7. **Limitations and Future**

INTRODUCTION

- Host Organisation
- Group Members
- Prologue

Host Organisation

- Lensell
  - AI-powered platform
  - Helps Australian companies on the financial markets
  - Helps investors and analysts access performance data

- Founder & CEO of Lensell: Dr Laura RUSU
  - Fintech CEO
  - Passion researcher
  - Publisher author

LENSELL ®

Group Members

Andrew STRINGER
694968

Team Leader
Data Scientist

Chao JIA
958973

Lead Researcher
Data Analyst

Wei LI
956833

Project Manager
Data Scientist

Xin WEI
980496

Client Contactor
Data Analyst

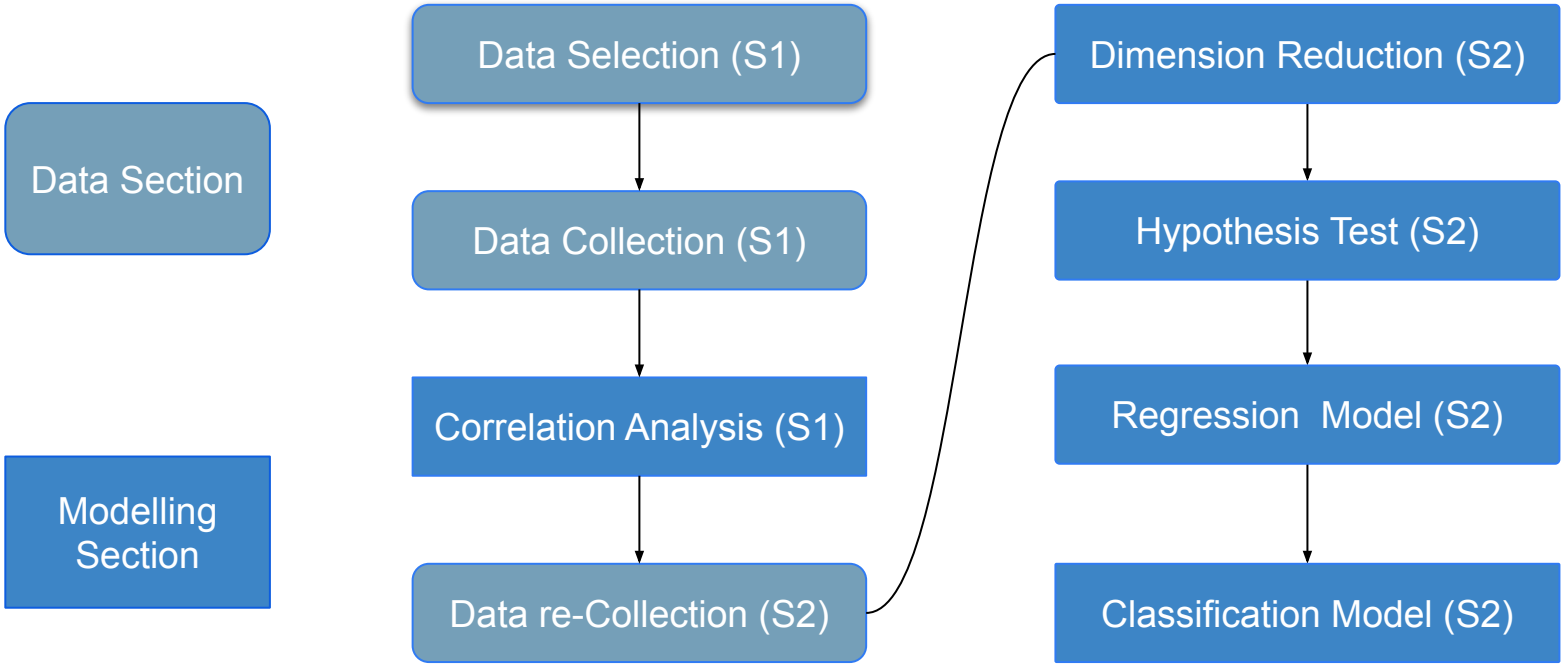
Prologue

- Topic:
  - Investigate the relationship between climate events and corporate financial performance
- Background:
  - Climate events impact the financial markets
  - Limited discussions and researches
- Scope of Project:
  - Data Science & Finance
  - Hypothesis Test
  - Statistical Methods
  - Machine Learnign Algorithms

PROJECT OVERVIEW

- **Stages of Pipeline**
- **Challenges**

Stage of Pipeline



Challenges

- Data
  - Data Security
  - Limited Data Points
  - Data Cleaning
  - Joining datasets
- Model
  - Statistical Model Selection
  - Predictors Dimension Reduction
  - Lack of Previous Work Investigating this Relationship
  - Hypothesis Testing

DATASET DISCUSSION

- Financial Datasets
- Climate Datasets
- Joining Datasets

Financial Data Collection

- Issues from Semester 1:
  - Data quality from Yahoo Finance
  - Messy agriculture industry divide
- Solutions:
  - Data acquisition through XBRL API from U.S. Securities and Exchange Commission (SEC) Website
  - Standard Industrial Classification (SIC) Code

Financial Data Collection

- Issues from SEC website:
  - Limited returned value from XBRL API
  - Limited API valid companies
- Solutions:
  - Manually collect Financial Data from SEC website
  - Combine of Sub-Industries (totally 23 valid companies):
    - Agricultual production
    - agricultural livestock
    - Agricultural service
    - agricultural chemicals

Financial Ratios Selection

- Corporate Financial Performance (CFP) measurement Selection:
  - ROA
  - ROE
  - Current Ratios
  - Net Margin
  - Operating Margin

Climate Datasets

- Climate Data Selection
  - Temperature Data:
    - Berkeley Earth.org
    - Access more lastest data
    - Yearly and monthly average temperature (2002-2020)
  - Bushfire Data:
    - National Interagency Fire Center. gov
    - Indicator selection: Number of fires V.S. Burnt areas
    - Yearly burnt areas by U.S. states (2002-2020)

Joining Datasets

- Challenges:
  - CFP reporting periods have various start months
  - Temp data connected by month within reporting period
  - Bushfire data by year (Jan. -- Dec.)
- Handling bushfire data:
  - Majority of companies report Jan. -- Dec.
  - Impact delay assumed, majority of fires being in U.S. summer period assumed

LITERATURE / TECHNIQUES REVIEW

- **Related Literature Areas**
- **Suitability of Approaches to Our Work**

LITERATURE / TECHNIQUES REVIEW

- Approaches taken in related literature areas
  - Climate-finance lit areas
    - distinguished by climate focus and financial performance measures
  - Climate change risk--industry:
    - Linear modelling
  - Temperature--stock market:
    - Risk modelling (e.g. ΔCoVaR)
    - Quantile regression
  - Natural disaster--stock returns:
    - GARCH modelling

LITERATURE / TECHNIQUES REVIEW

- Suitability of approaches
  - Most suitable:
    - Climate change risk--industry and linear regression
  - Challenge with other methods:
    - Models make use of high frequency of stock market data
    - Quantile regression appears better suited to studying multiple companies together

METHODOLOGY

- **Feature Engineering**
- **Regression Model & Hypothesis**
- **Classification Model**

Hypothesis

- Competing Hypothesis

  H_0: Climate factors do not impact corporate financial performance.

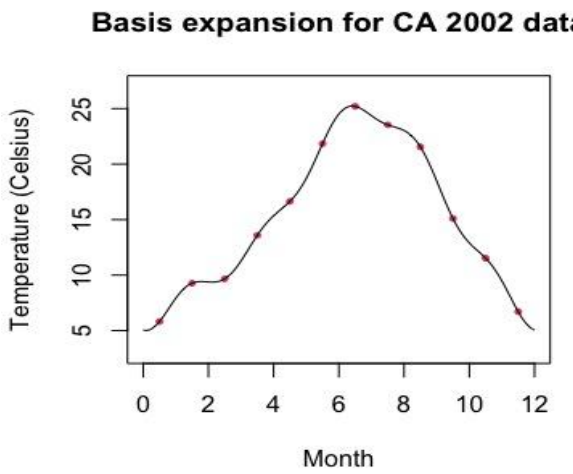
  H_a: Climate factors impact corporate financial performance

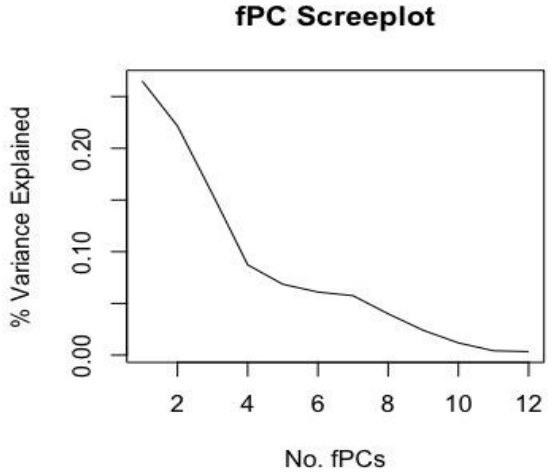
Methodology

- Aims and Methodology of Data Analysis
  - Identify plausible climate--cfp relationships
    - Dimension reduction of monthly temp data
    - Use linear and nonlinear methods
  - Test whether found relationship is statistically significant
    - Using LOOCV
    - Using permutation test
  - Assess study's collective results set for false positives
    - Using binomial test

Functional Principle Component Analysis

- Functional PCA for temperature dimension reduction
  - Representation in Fourier basis
  - How many functional PCs to use?





Independent Component Analysis

- Method: decompose data into the independent components
- Idea: one season corresponds to one component
- Comparison:
  - independence instead variance;
  - non-Gaussian assumption over data;
  - without orthogonality assumption over components.

Regression Model

- Model: $CFP \sim temperature(fPCs) + bushfires$
- Relationship chosen by maximising adjusted R^2
  - However, this makes t-test results unreliable
- H_0: expect to see identified model with a similar adjusted R^2 value given that there is no relationship
- Permutation test used to construct distribution of adjusted R^2 given H_0

Classification Model

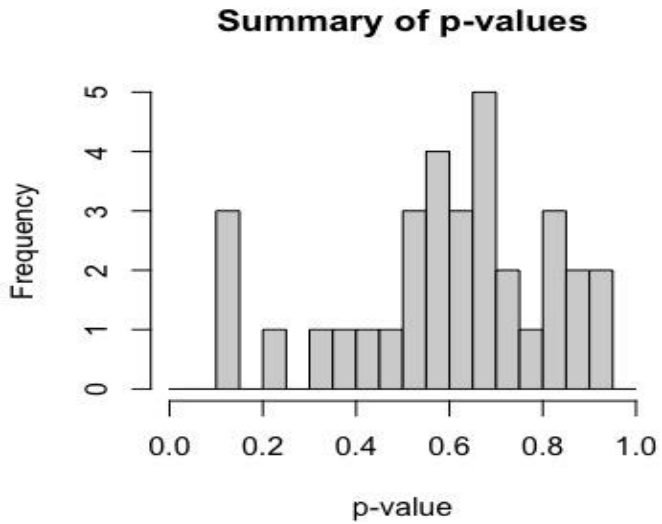
- Logistic regression:
  - Model: $\Delta\, CFP \sim LR(\, temperature(\, f\, PCs) \,+\, bushf\, ires)$
  - Idea: use activate function to transform features
- Support vector machines (kernel trick):
  - Model: $\Delta\, CFP \sim SVM(\, temperature(\, f\, PCs) \,+\, bushf\, ires)$
  - Idea: map the features into high dimessional kernel space
- Random forest:
  - Model: $\Delta\, CFP \sim RF(\, temperature(\, f\, PCs) \,+\, bushf\, ires)$
  - Idea: approximate non-linear relationship via piecewise constants

THE UNIVERSITY OF MELBOURNE

RESULTS & CONCLUSION

Regression & Hypothesis

- Insufficient evidence to reject null hypothesis for any relationship
  - Permutation test at 0.05 level
  - min p-value of 0.124
  - max p-value of 0.948



THE UNIVERSITY OF MELBOURNE

Classification Results

Model	Logistic Regression		Support Vector Machine		Random Forest	
Reducation	**ICA**	**FPCA**	**ICA**	**FPCA**	**ICA**	**FPCA**
CFP						
CurrentRatio	0.526	0.473	0.502	0.468	0.466	0.428
NetMargin	0.382	0.312	0.377	0.293	0.374	0.388
OperatingMargin	0.332	0.310	0.332	0.275	0.428	0.430
ROA	0.467	0.465	0.402	0.354	0.491	0.449
ROE	0.373	0.470	0.355	0.397	0.424	0.421
Avg.	0.416	0.406	0.394	0.357	0.436	0.423

CONCLUSION

- Correlation analysis points a fragile relationship over some comps.
- Evidences are not enough to reject the null hypothesis using linear regression
- Classification shows an acceptable results
- In sum, financial performance displays a weak relationship with climate data

LIMITATIONS & FUTURE

- Limited information in each company
- Limited granularity in climate data
- Invetisgate causal relationship
- Focus on specific mechanisms rather than broader CFP impacts

THE UNIVERSITY OF
MELBOURNE

Thank you