FIT1043 Introduction to Data Science

Module 5
Data Analysis Process Case Studies

Lecture 10 – Part II

Background: Different Data Types

- Spatial data (also known as geospatial data):
  - is information about a physical object that can be represented by numerical values in a geographic coordinate system.
- Temporal data:
  - is data that varies over time.
- Spatio-Temporal data:
  - is the integration of space and time.

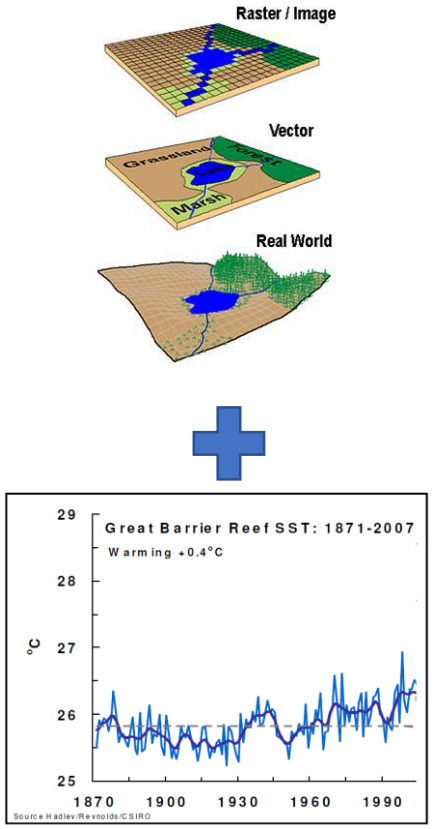
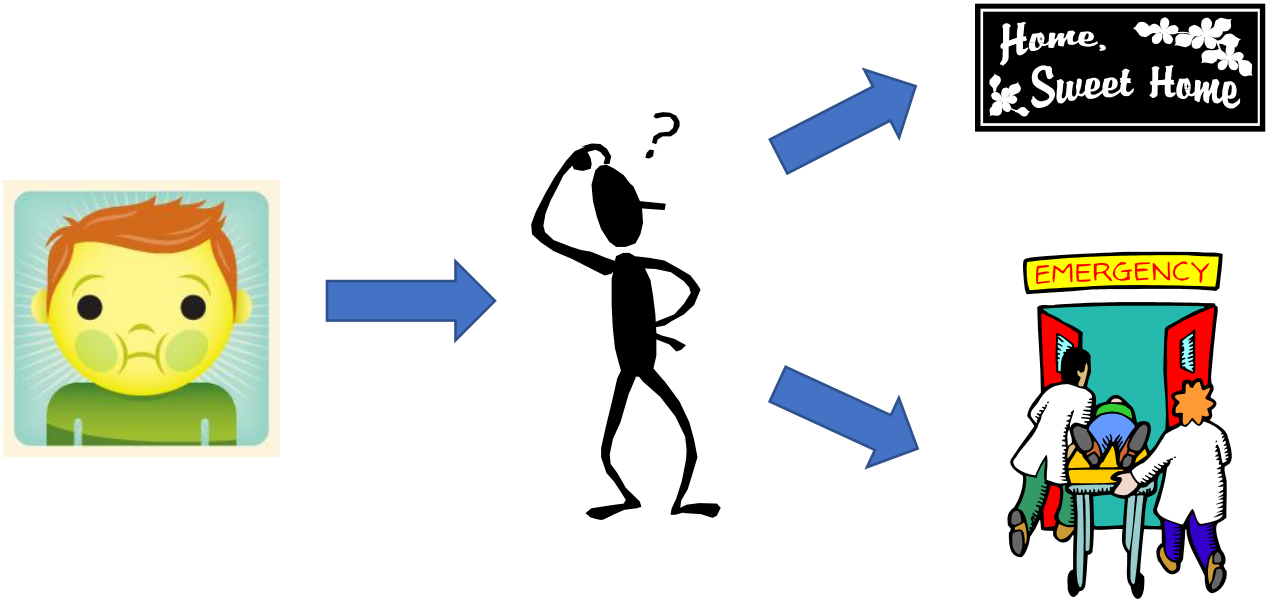


Image planet.botany.uwc.ac.za

Data Analysis Case Studies
Diagnosis of Pneumonia

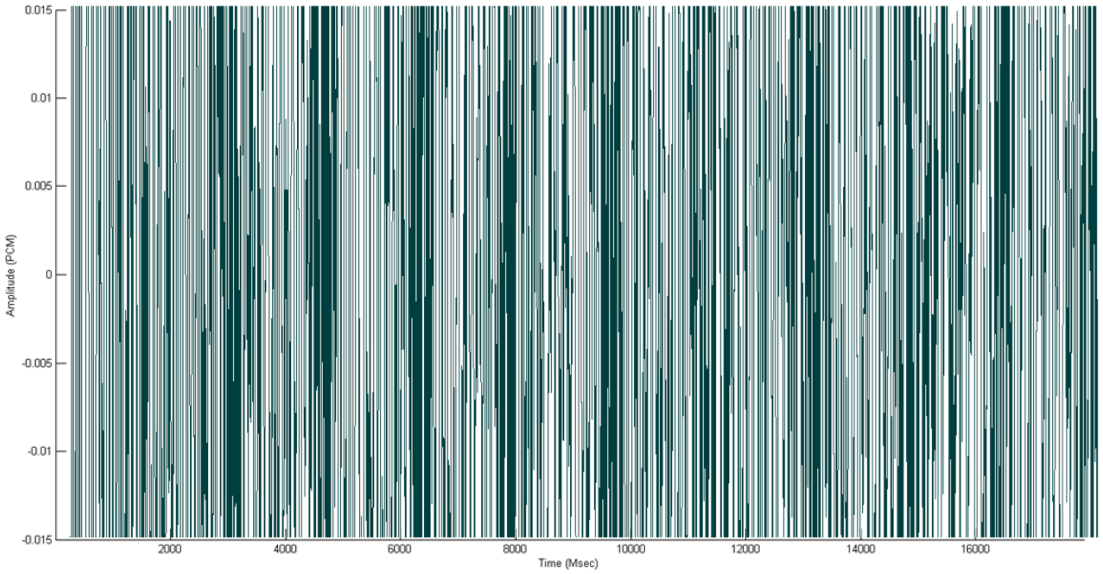
Diagnosing pneumonia is difficult

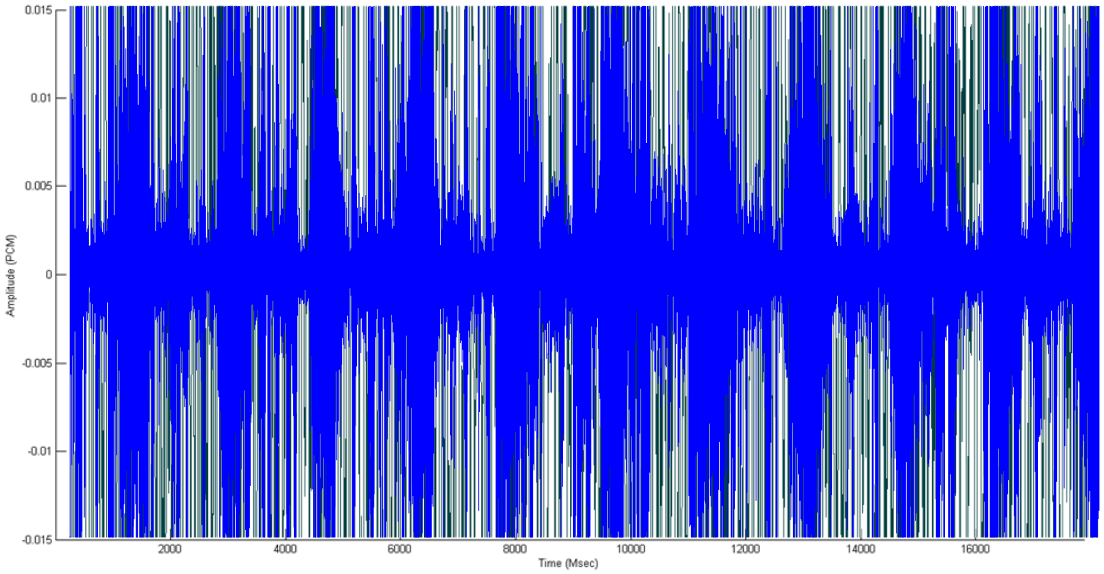
- Pneumonia is the single biggest killer of children.
- It kills 1.5 million children under 5 per year, mostly in developing countries.
- The key to change this lies in early diagnosis.

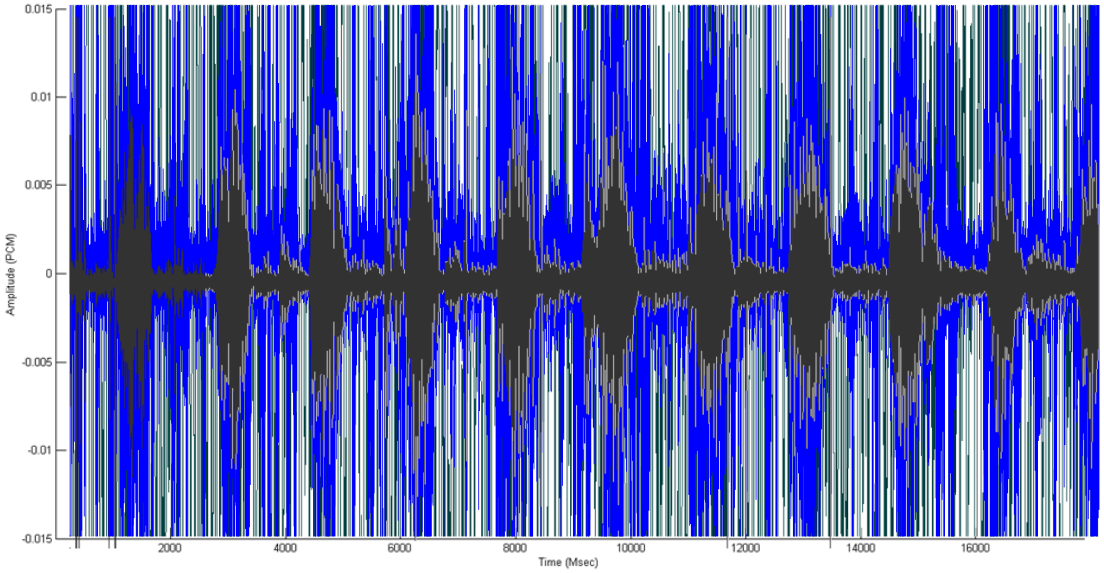


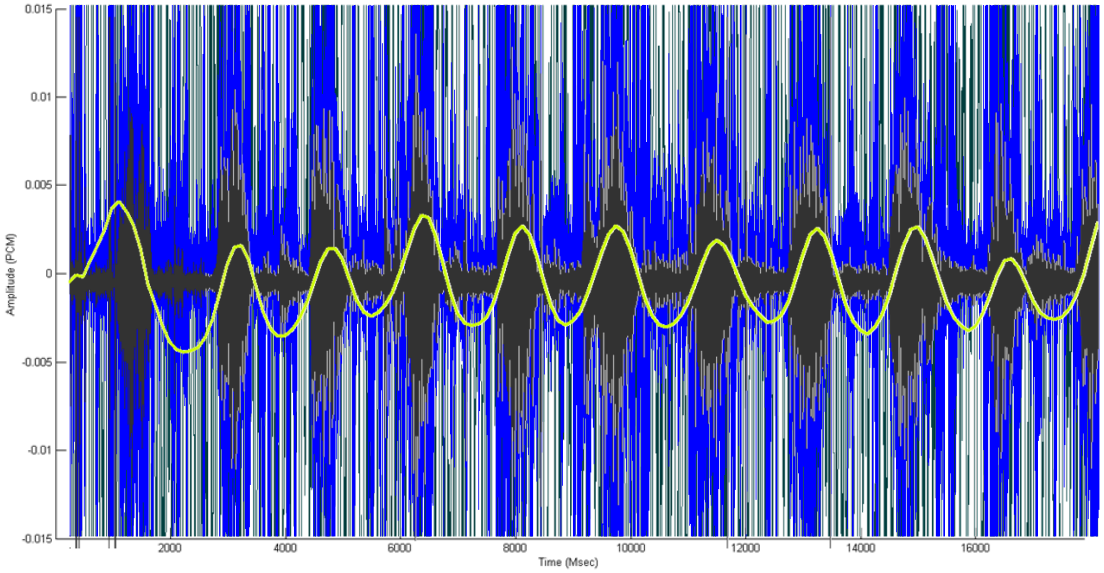
IoT based solution: StethoCloud

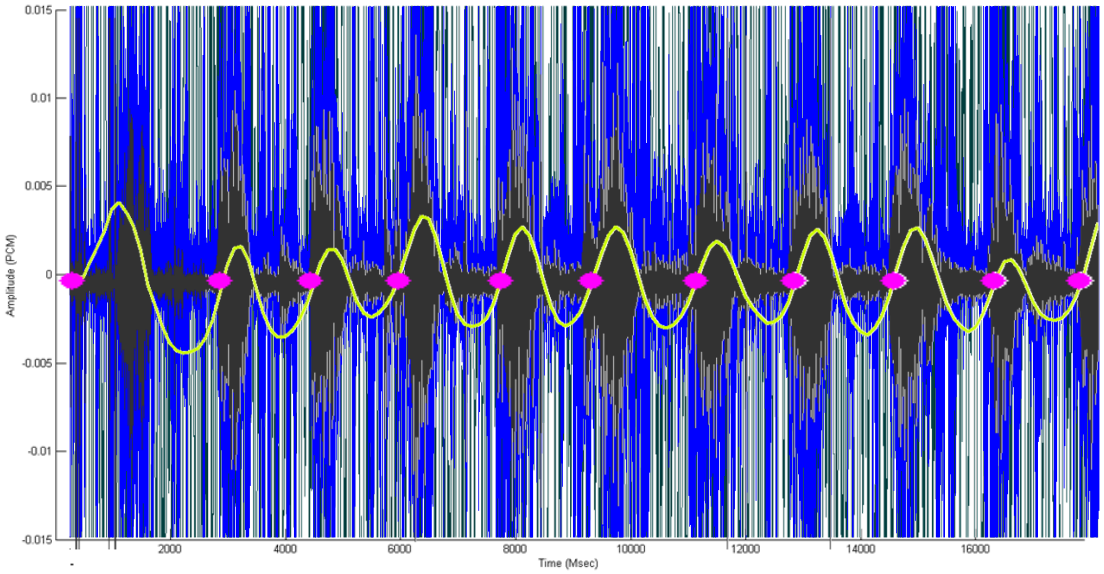
Respiratory Data is collected from six different locations


0.015
0.01
0.005
0
-0.005
-0.01
-0.015
Amplitude (PCM)
2000
4000
6000
8000
10000
12000
14000
16000
Time (Msec)


0.015
0.01
0.005
0
-0.005
-0.01
-0.015
Amplitude (PCM)
2000
4000
6000
8000
10000
12000
14000
16000
Time (Msec)

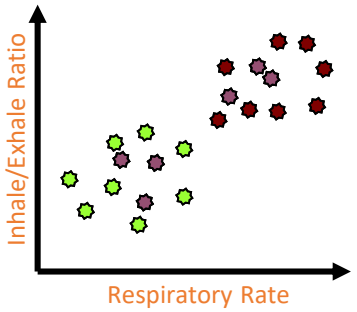

0.015
0.01
0.005
0
-0.005
-0.01
-0.015
Amplitude (PCM)
2000
4000
6000
8000
10000
12000
14000
16000
Time (Msec)


0.015
0.01
0.005
0
-0.005
-0.01
-0.015
Amplitude (PCM)
2000
4000
6000
8000
10000
12000
14000
16000
Time (Msec)


0.015
0.01
0.005
0
-0.005
-0.01
-0.015
Amplitude (PCM)
2000
4000
6000
8000
10000
12000
14000
16000
Time (Msec)

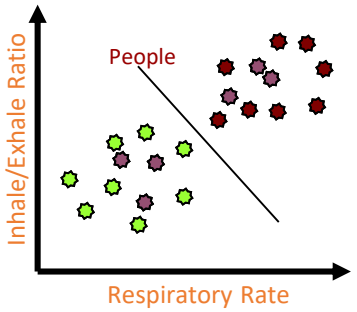
Classification and Clustering

- Classification
  - Have labels for some points



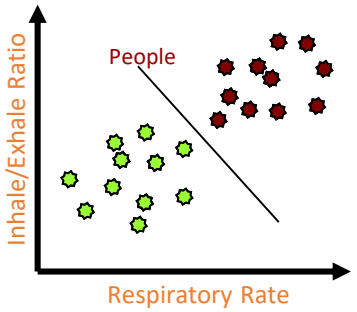
Classification and Clustering

- Classification
  - Have labels for some points
  - Want a "rule" that will accurately assign labels to new points



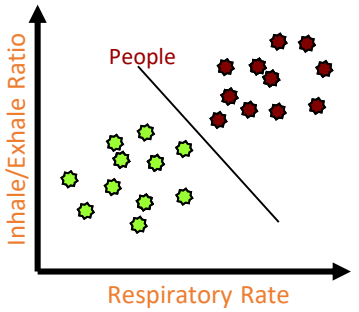
Classification and Clustering

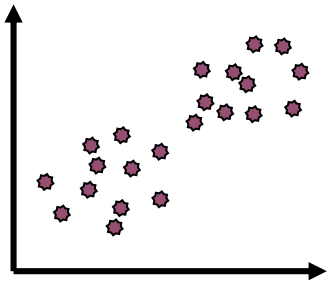
- Classification
  - Have labels for some points
  - Want a "rule" that will accurately assign labels to new points



Classification and Clustering

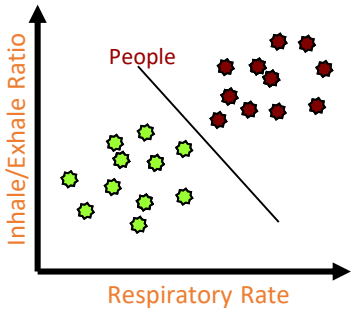
- Classification
  - Have labels for some points
  - Want a "rule" that will accurately assign labels to new points
- Clustering
  - No labels



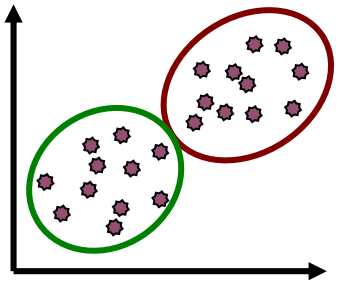


Classification and Clustering

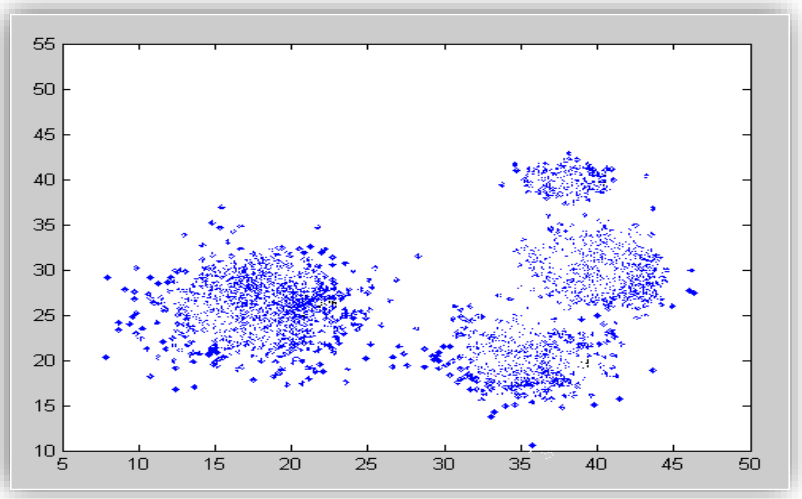
- Classification
  - Have labels for some points
  - Want a "rule" that will accurately assign labels to new points



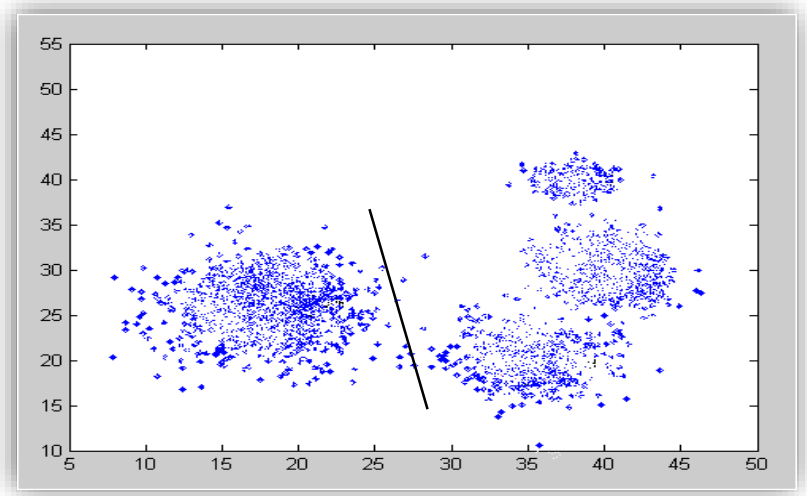
- Clustering
  - No labels
  - Group points into clusters based on how "near" they are to one another
  - Identify structure in data



This is what the data might look like



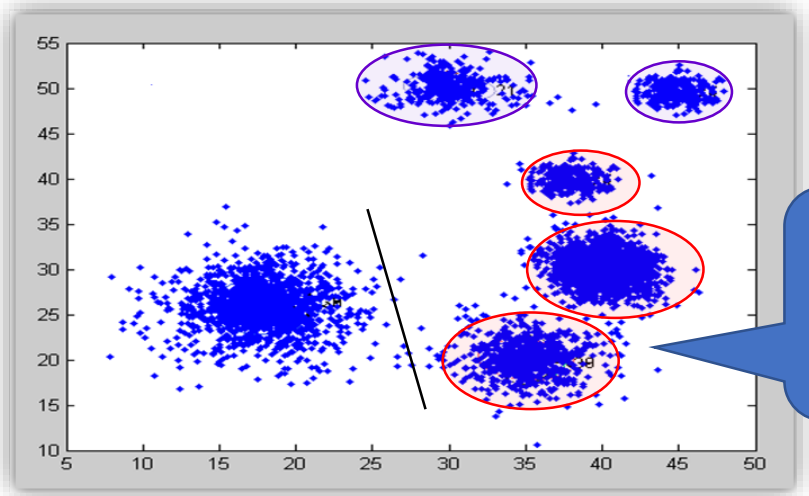
This is what the data might look like



Normal Subject **Pneumonia**

Classification

This is what the data might look like



Clustering

Possible patterns emerging with more data, e.g., different subtypes

Normal Subject

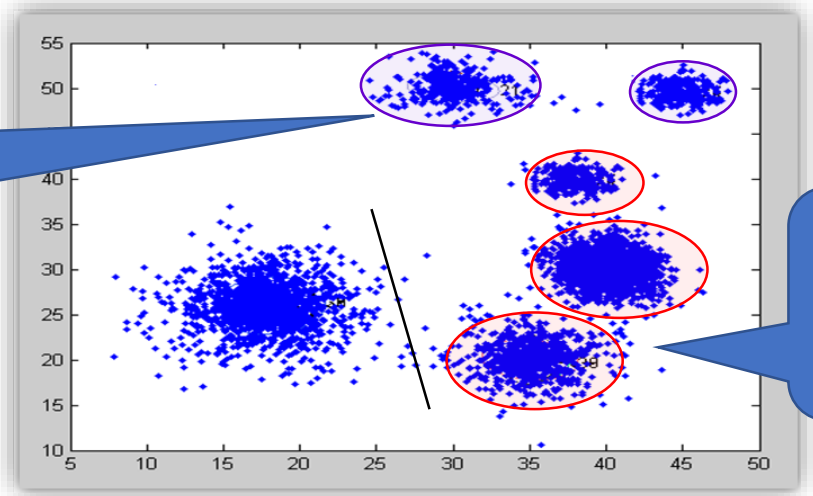
Pneumonia

Classification

This is what the data might look like

Other Diseases

This technique can directly be used for other diseases, e.g. asthma



Clustering

Possible patterns emerging with more data, e.g., different subtypes

Normal Subject

Pneumonia

Classification

Data Analysis Case Studies
Car-Racing Driver Distraction Detection

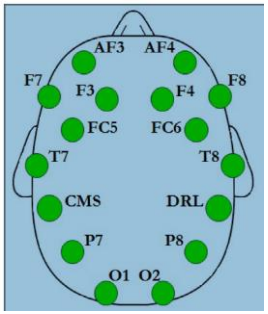
Image emotiv.com

Why is it important?

- Monitoring driver attention has a direct effect on decreasing injury/fatality rates, and improving his performance.
- In **car racing** environments reaction times are short, and distraction leads to a reduction in driver's performance during a race.

Brain Computer Interface (BCI) Approach

Emotiv EPOC headset



14 Brain EEG Signals
+ Gyroscope

Experimental Setup

- Simulation

- Talking to passenger
- Mobile call, Recording call
- Solving simple mental arithmetic challenges placed on the road, assessed by the driving through the chosen answer from a range of alternatives on screen

EEG data analysis

- Classification
  - Random forest has best results
  - 73.5% detection accuracy with EEG
  - 81% detection accuracy with EEG + Gyroscope

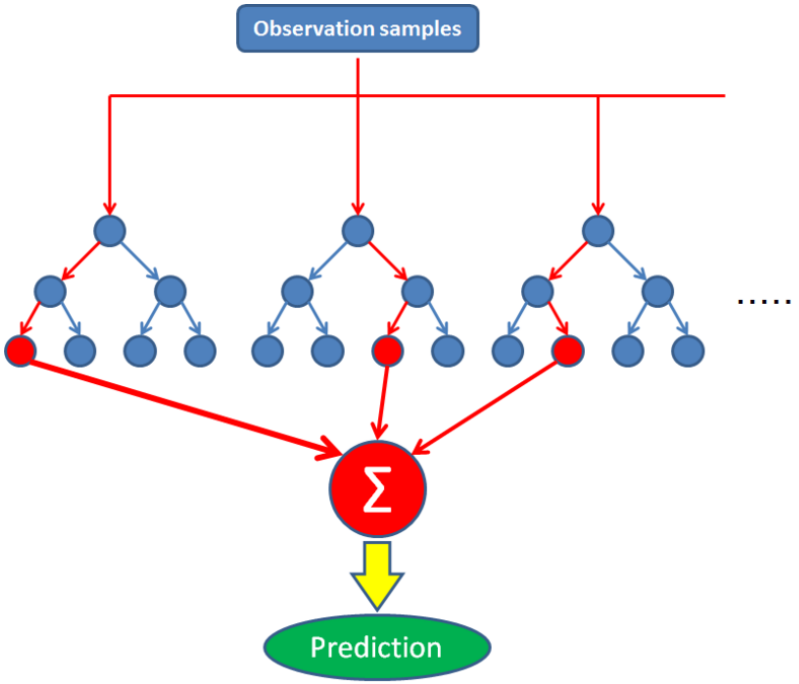


Image cdn-ak.f.st-hatena.com

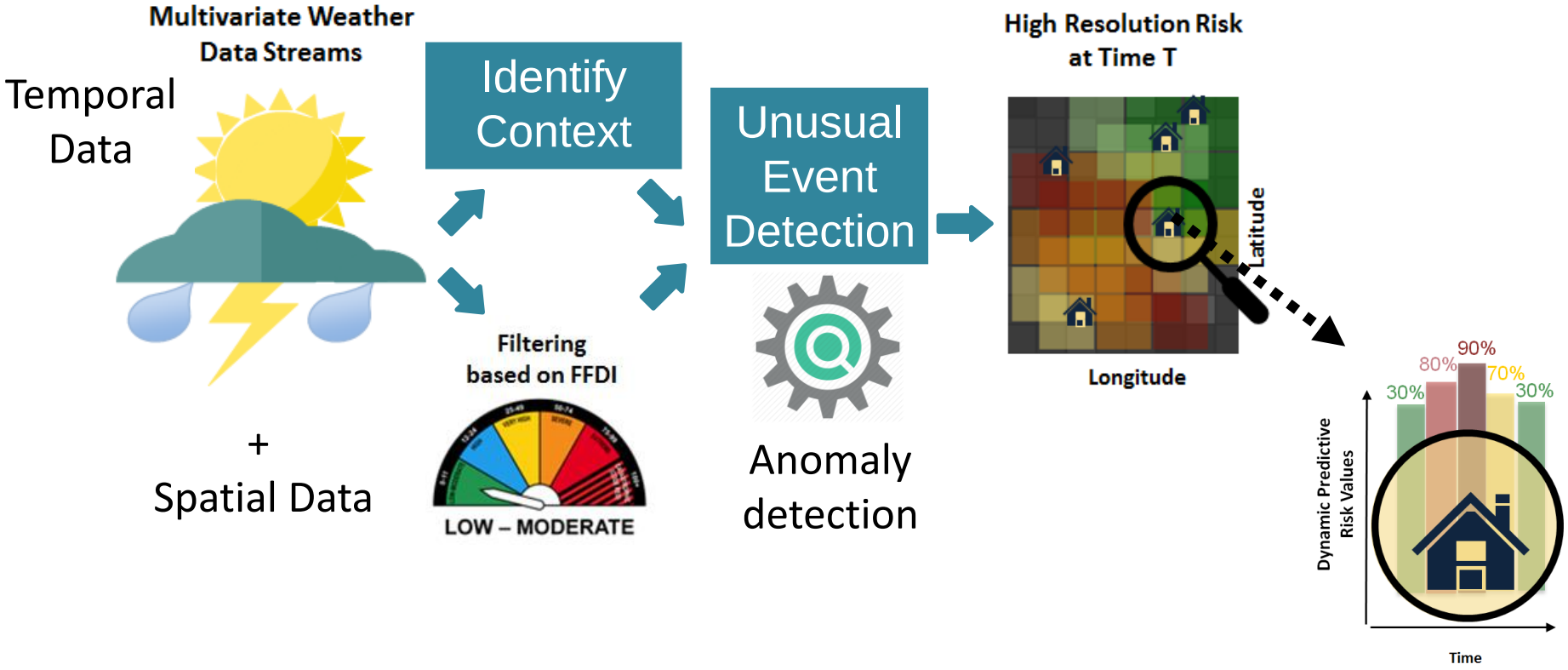
Data Analysis Case Studies
Disaster Management

Image abc.net.au

Why bushfires?

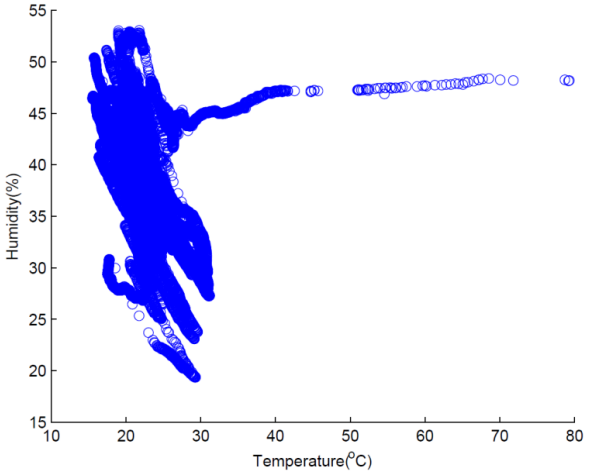
- Bushfires have shaped the Australian landscape for many years. With urban expansion and changing weather patterns the thread to homes and families is increasing.
- Australia experiences bushfires as the most damaging disasters.
- Examples:
  - Ash Wednesday bushfires 1983
  - Black Saturday 2009
- Ability to ***predict*** the risk of bushfires is crucial in helping emergency services in their decision-making processes, to mitigate and reduce the impact of such events.

Dynamic bushfire risk prediction



Anomaly Detection

- Anomaly: data points that are inconsistent to the normal data points
- Anomaly detection: the process of finding anomalous patterns in data sets
- In this application, anomalies are relevant to the episodes of time with high bushfire risk

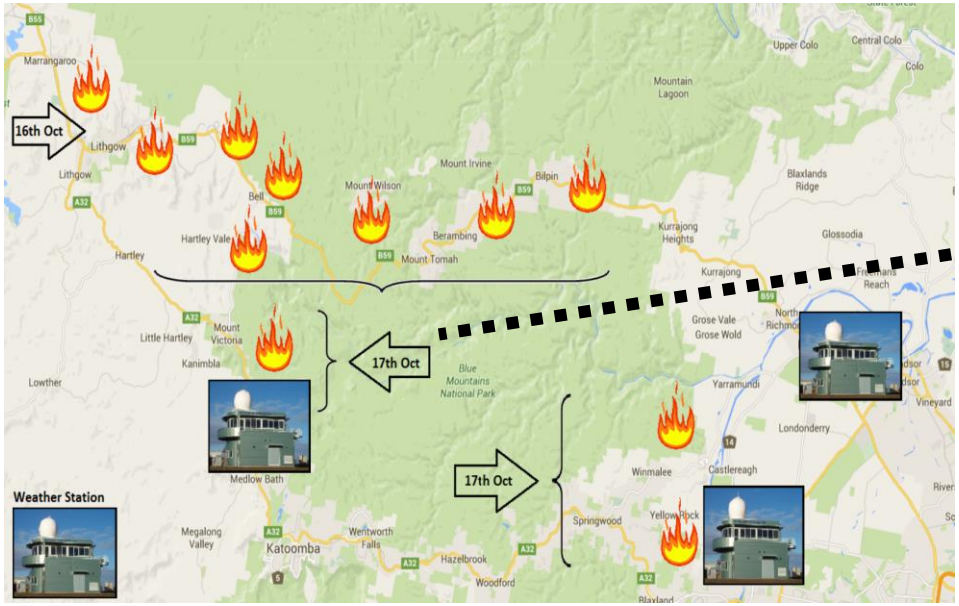


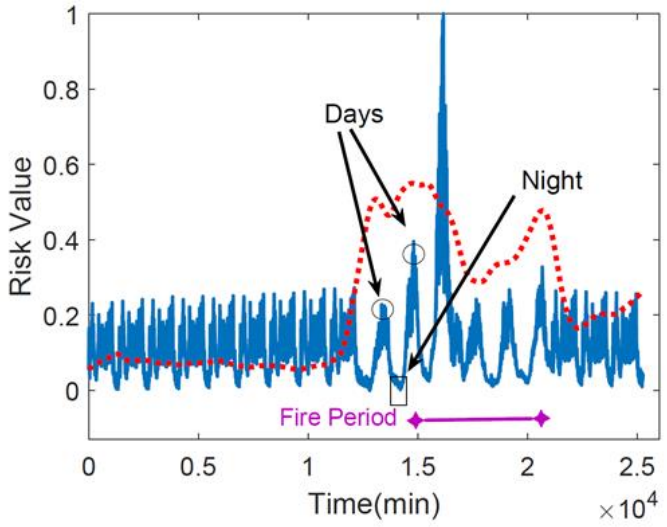
Case study- 2013 Blue Mountains Fire, NSW

- Why Blue Mountains?
  - 11 days of intense bushfires
  - 2 fatalities
  - 248 houses destroyed
  - 183.4M AUD in insurance claims

- Data Available
  - ~ 45 thousands houses
  - historical weather measurement

Case study- 2013 Blue Mountains Fire, NSW



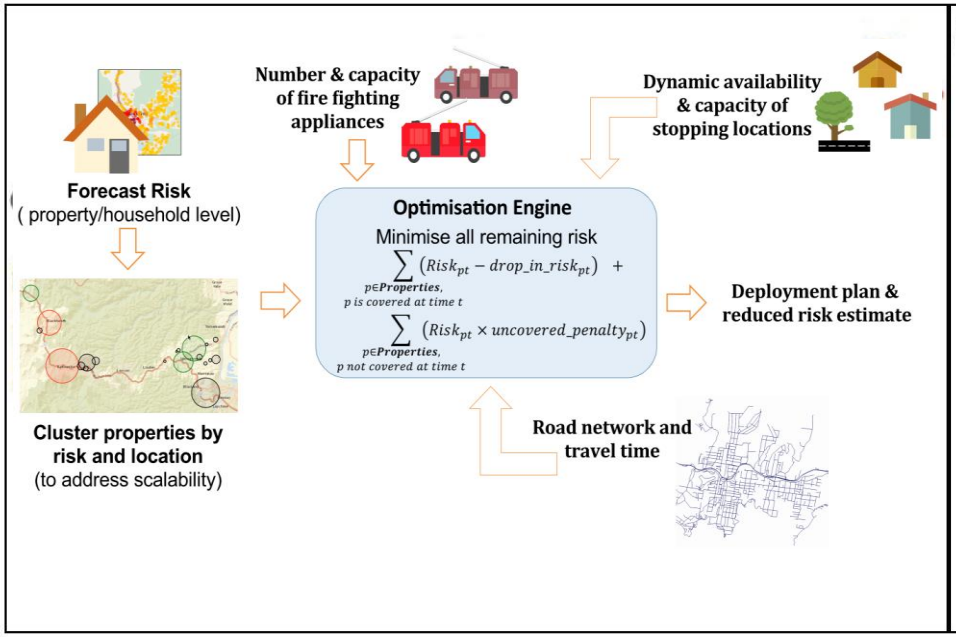


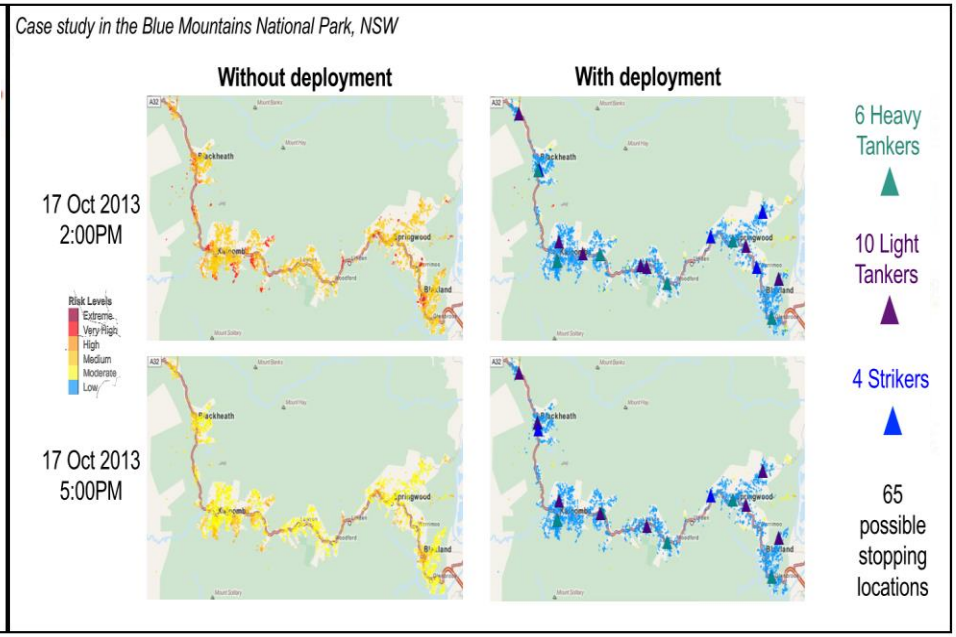
Our Method Risk Values

FFDI Risk Values

Fire Fighting Appliance Pre-Deployment

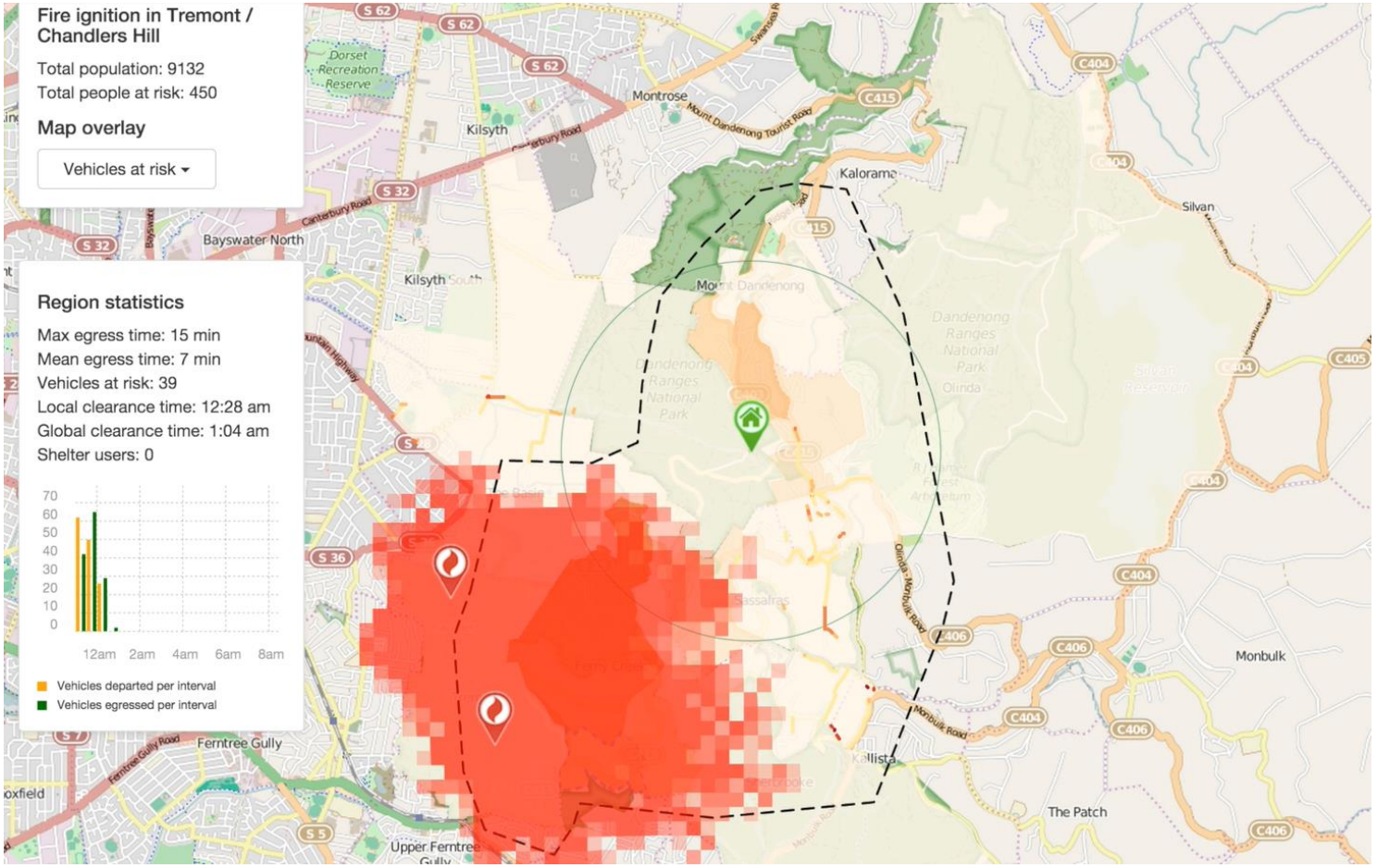
- Optimise locations of available fire fighting resources subject to forecast risk and roadside stopping locations with time dependent availability.





Evacuation Planner

- Spatio-Temporal fire expansion



Unit Schedule: Next Week

Module	Week	Content
1	1 2	Overview and look at projects (Job) roles, and the impact
2	3	Data business models / application areas
3	4 5	Characterising data and "big" data Data sources and case studies
4	6 7	Resources and standards Resources case studies
5	8 9 10	Data analysis theory Regression and decision trees Data analysis process
6	11 12	**Issues in data management** GUEST SPEAKER & EXAM INFO.