cenit
Science for Transport

September 23, 2022

bin
Behavior in Networks Studies Unit

Fundamentals and Applications of Weak Learners

Muhammad Awais Shafique

Centre d'Innovació del Transport (CENIT)
Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain

Introduction

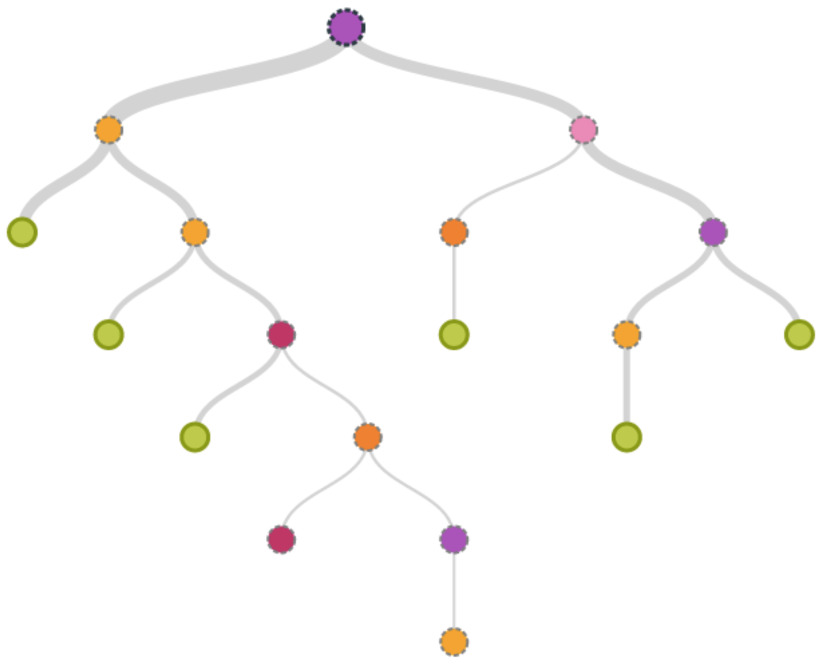
- When predicting, the least one can do is Random Guessing

- **Weak Learner**
- "A weak learner produces a classifier which is only slightly more accurate than random classification."

Pattern Classification Using Ensemble Methods, pg 21, 2010

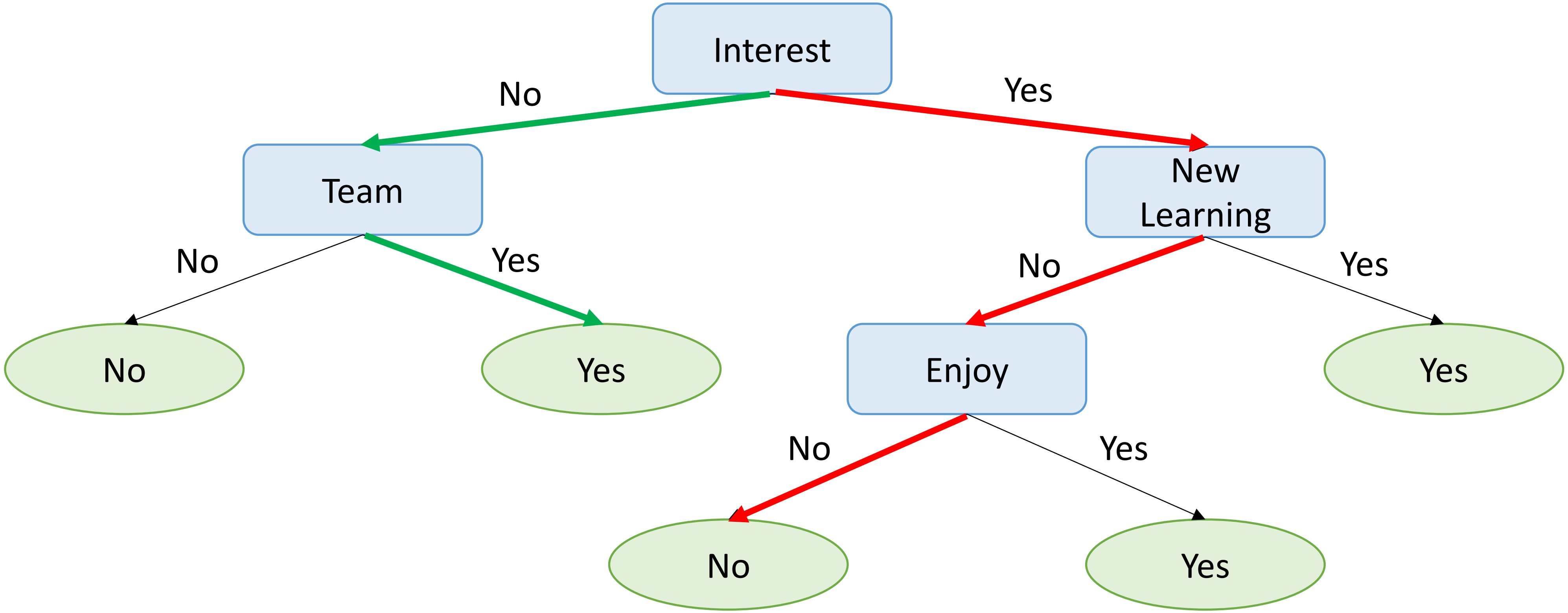
Introduction

- A popular example is Decision Tree.
- Weakness can be controlled by the depth of tree.
- Weakest tree: only one node and binary decision made on only one variable.



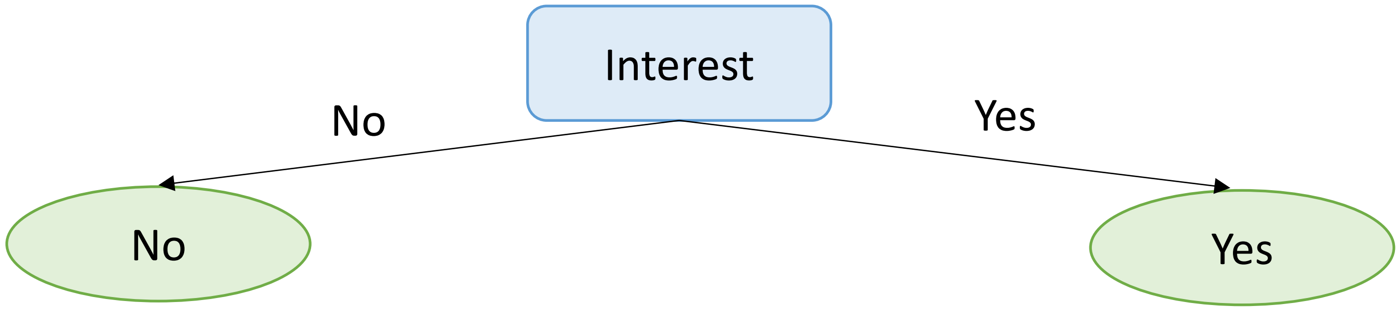
Introduction

- Example: Your decision to participate in the summer school.



Introduction

- Example: Your decision to participate in the summer school.



Introduction

- **Strong Learner**
- A strong learner produces a classifier that achieves arbitrarily good accuracy, better than random guessing.

- For modeling tasks, we aim to develop a strong classifier that makes predictions with good accuracy with high confidence.
- For instance, applying Support Vector Machines directly to the dataset.

Introduction

- In short

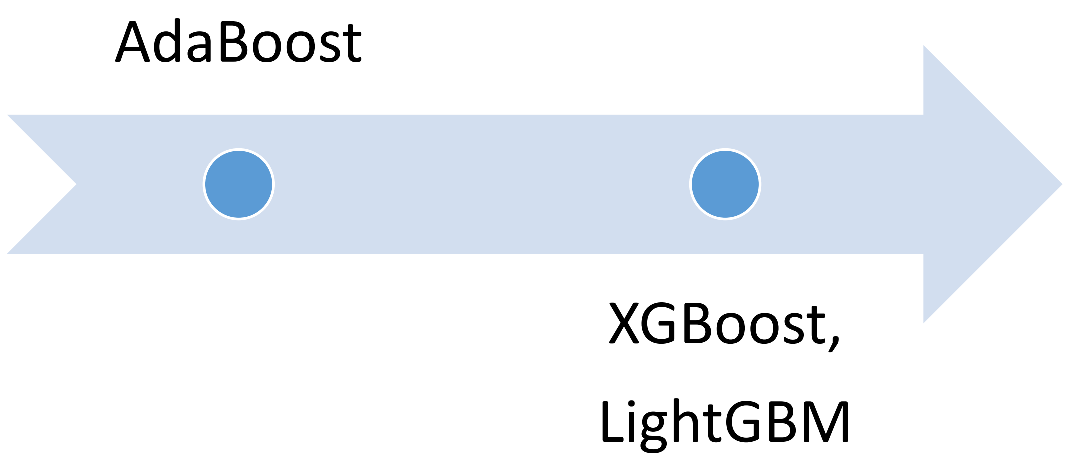
- Weak learners: Slightly better than random.
- Strong learners: Having good or even near-optimal accuracy.

- Are they equivalent?

YES

Boosting

- A strong learner can be constructed from many weak learners.
- This became the basis for boosting methods



Boosting

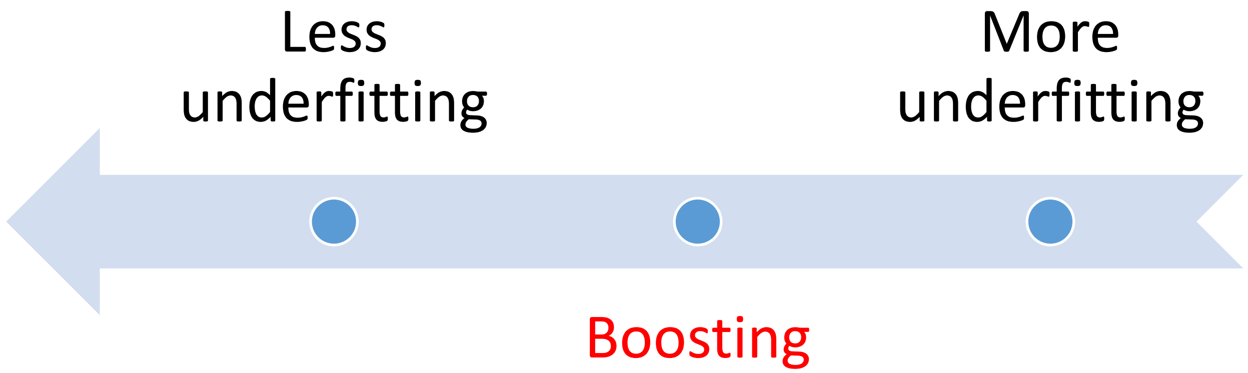
- Develop a large number of weak learners for a predictive learning problem.
- Combine them in a way to achieve a strong learner.

- Weak learners: Easy to prepare but not desirable.
- Strong learners: Hard to prepare and highly desirable.

Bagging vs. Boosting

Boosting

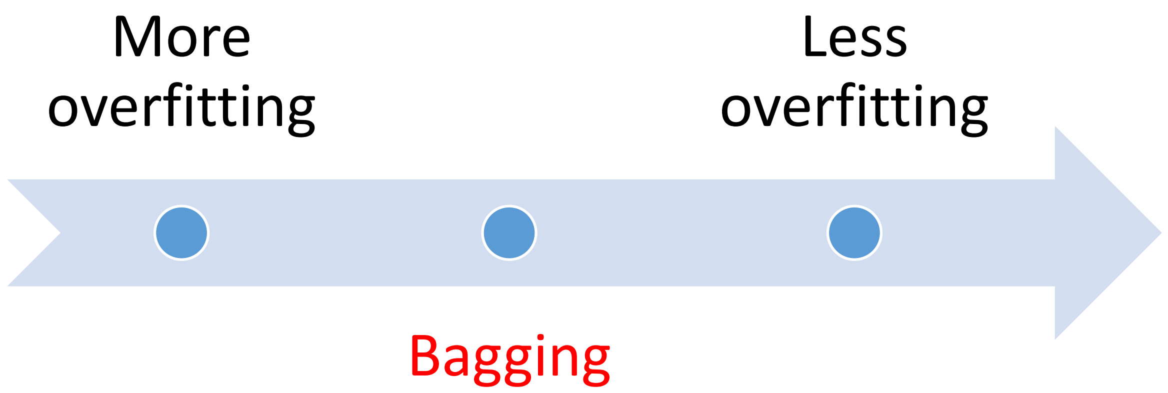
- Start with one decision tree stump (weak learner) and “focus” on the samples it got wrong.
- Train another decision tree stump that attempts to get these samples right.
- Repeat until a strong classifier is developed.



Bagging vs. Boosting

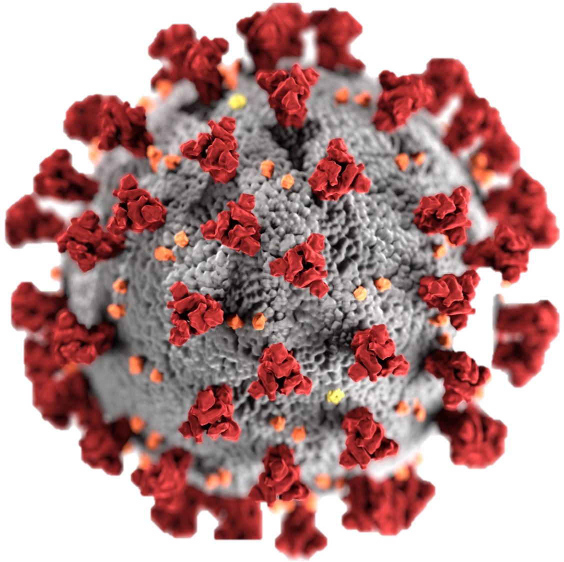
Bagging

- Train a number (ensemble) of decision trees from bootstrap samples of your training set.
- After the decision trees are trained, we can use them to classify new data via majority rule.

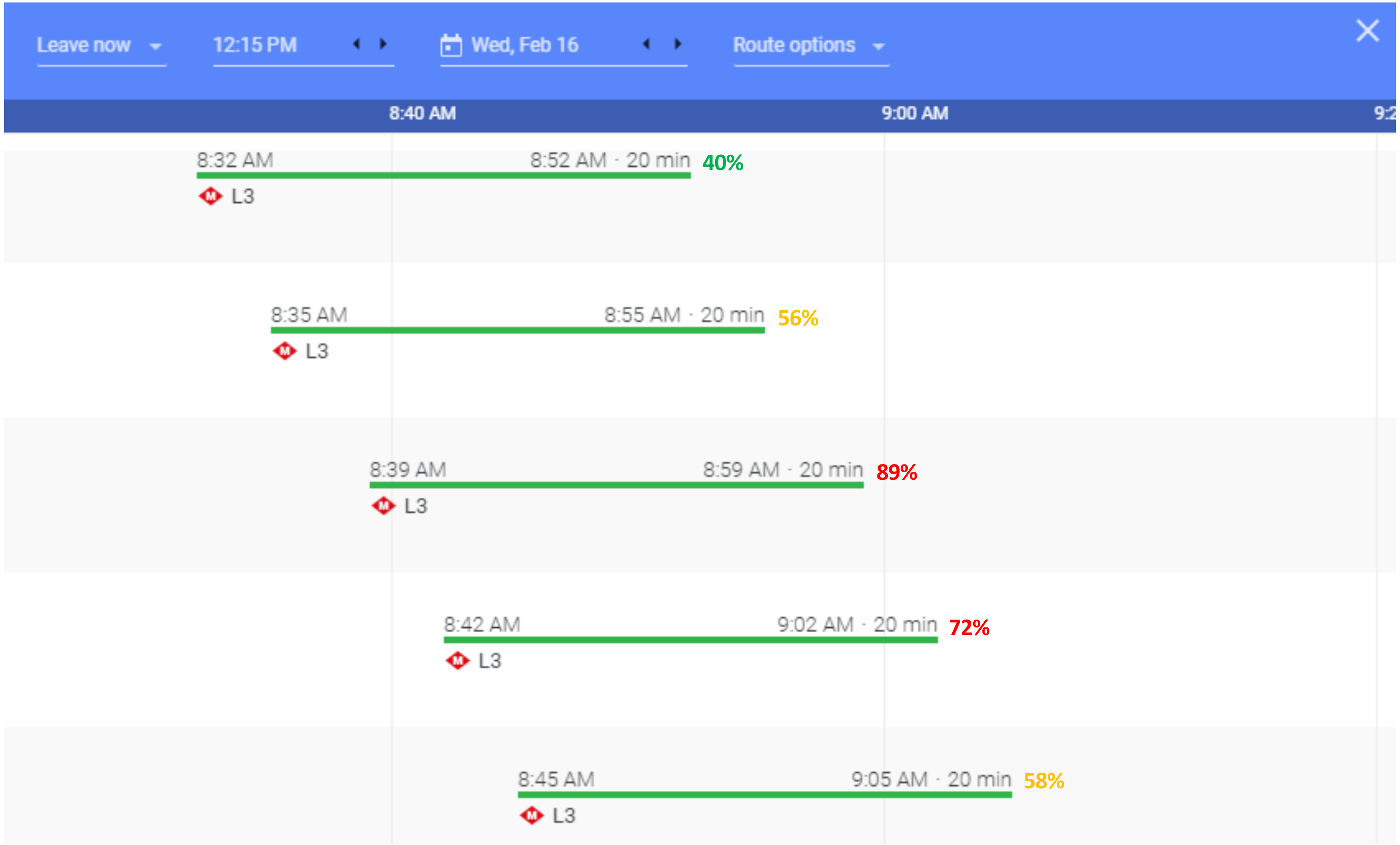


"Improving ridership by predicting train occupancy levels"

Consider a Scenario

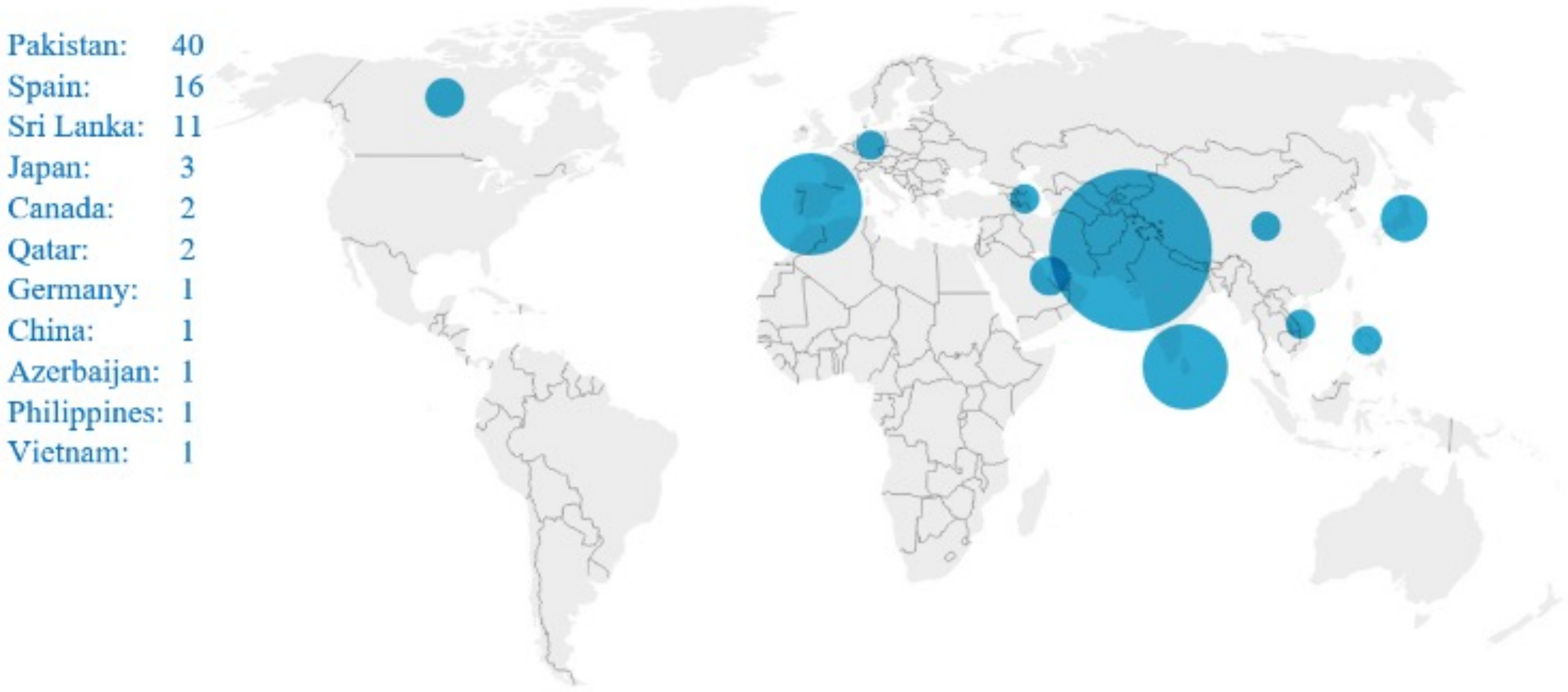



5303-5



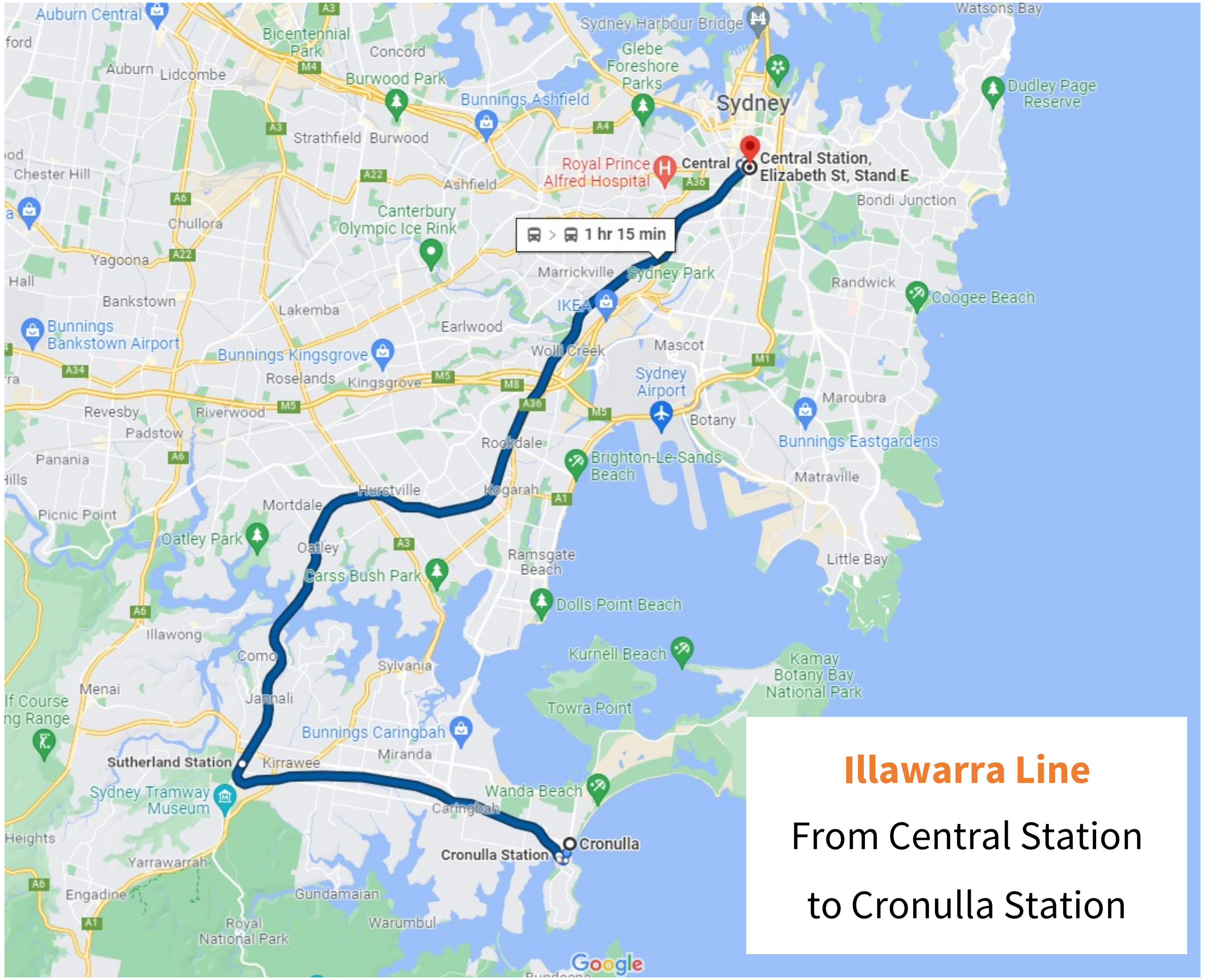
Leave now
12:15 PM
Wed, Feb 16
Route options
8:40 AM
9:00 AM
8:32 AM
8:52 AM · 20 min
40%
L3
8:35 AM
8:55 AM · 20 min
56%
L3
8:39 AM
8:59 AM · 20 min
89%
L3
8:42 AM
9:02 AM · 20 min
72%
L3
8:45 AM
9:05 AM · 20 min
58%
L3

Survey



- If future crowdedness levels are known 80% participants revealed that they will change their departure time and/or route to ensure less crowded transport.

Studied Train Route (NSW, Australia)

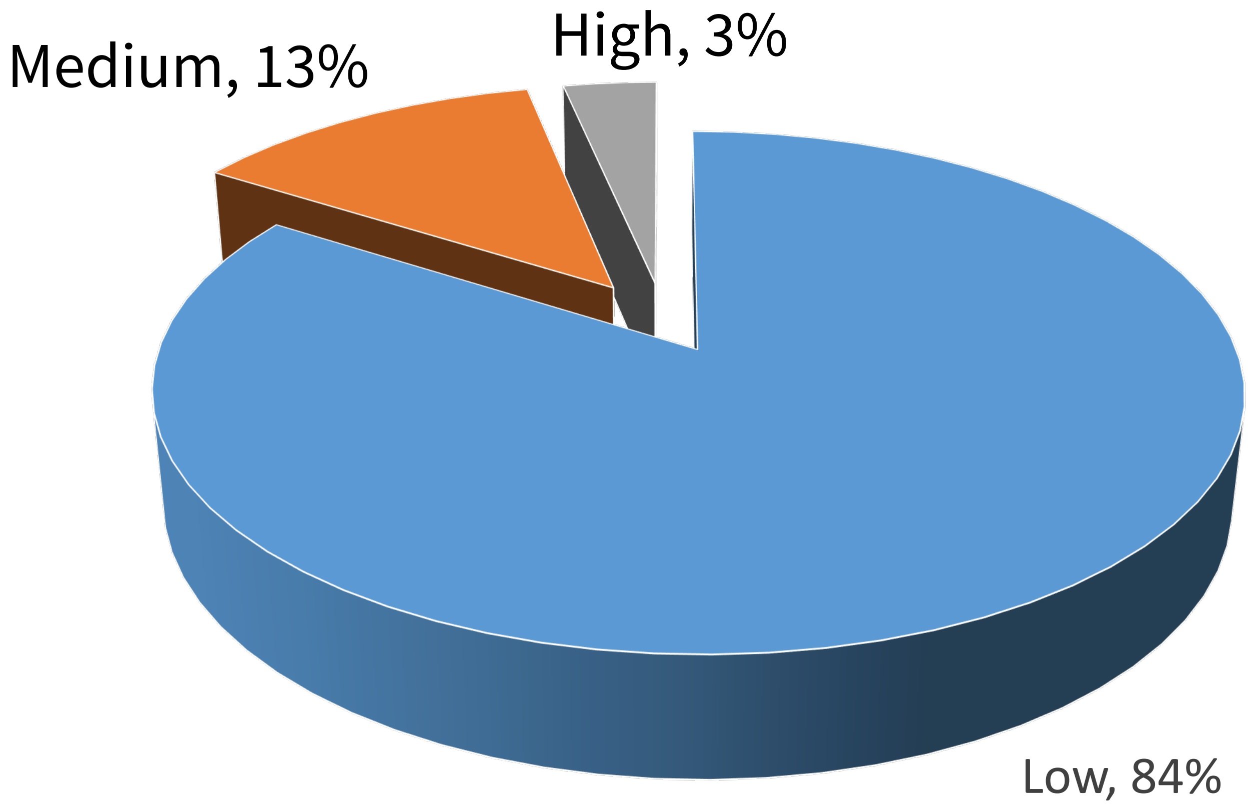


Illawarra Line

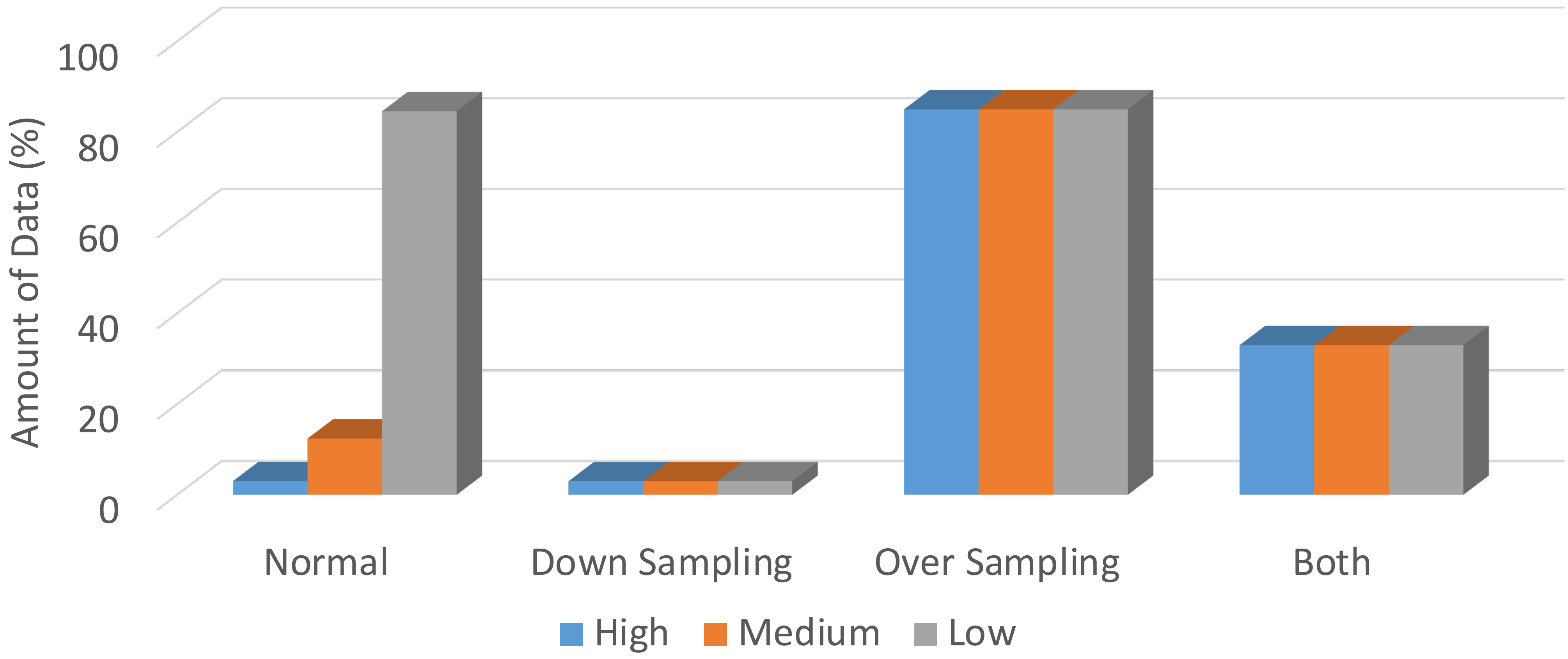
From Central Station

to Cronulla Station

Train Occupancy (Nov 2018 – Feb 2019)



Addressing Imbalanced Data



Overall Accuracy

Sampling	Classifier		
	XGB	RF	SVM
Normal	0.941	0.956	0.951
Down Sampling	0.930	0.940	0.935
Over Sampling	0.958	**0.959**	0.936
Both	0.955	0.955	0.939

Macro-averaged F1 Score

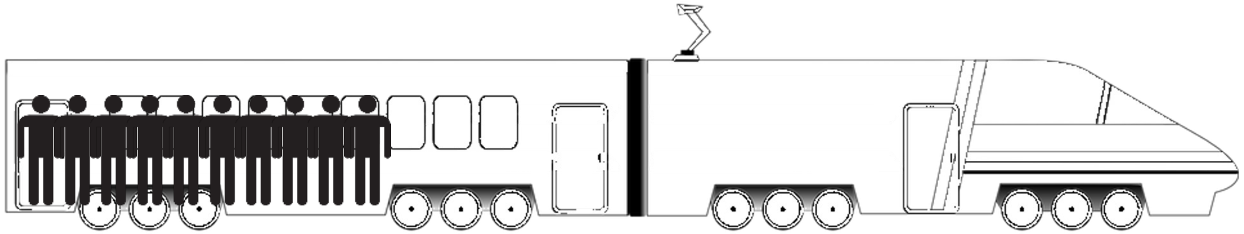
Sampling	Classifier		
	XGB	RF	SVM
Normal	0.824	0.876	0.865
Down Sampling	0.831	0.864	0.844
Over Sampling	**0.891**	0.883	0.826
Both	**0.891**	0.889	0.838

Challenges

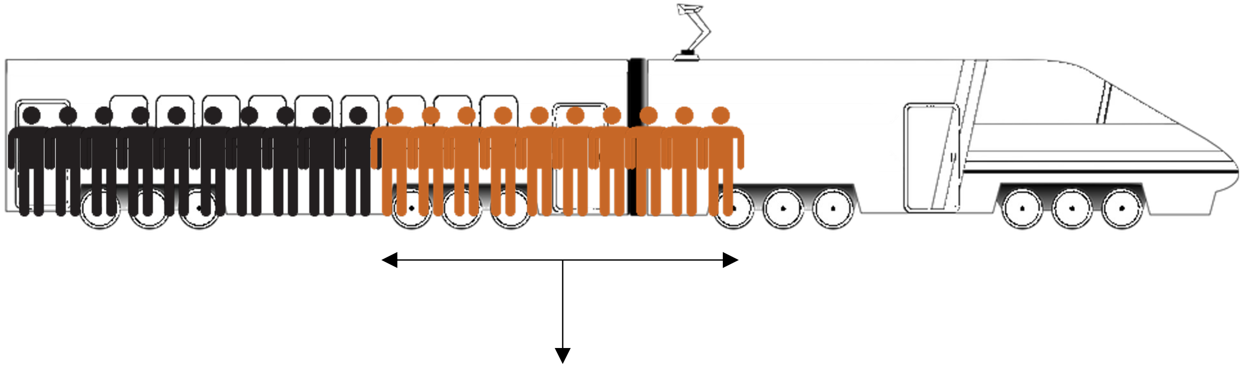
Imbalanced Data: Biased learning leading to skewed results.

Attracted Occupancy: Predicted Crowdedness values would be affected by changed travel behavior.

Predicted Crowdedness Level



Actual Crowdedness Level



Attracted Occupancy

cenit

Science for Transport

Centre d'Innovació del Transport (CENIT)
C/ Jordi Girona, 1-3, C3, S120, 08034, Barcelona
www.cenit.es

A research group of:

CIMNE

Annex

- $\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$

- $\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$

- $\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$

- $\text{F1 Score per class}, \text{F1}_c = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

- $\text{Macro} - \text{Averaged F1 Score} = \frac{\text{F1}_{\text{High}} + \text{F1}_{\text{Medium}} + \text{F1}_{\text{Low}}}{3}$