Predicting Himalayan Summit Success

Capstone Project 1

Jacques Poolman

Springboard

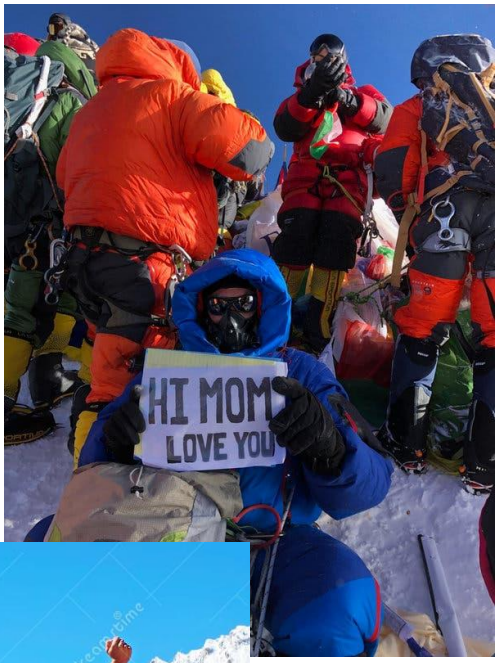
Predicting Himalayan Summit Success

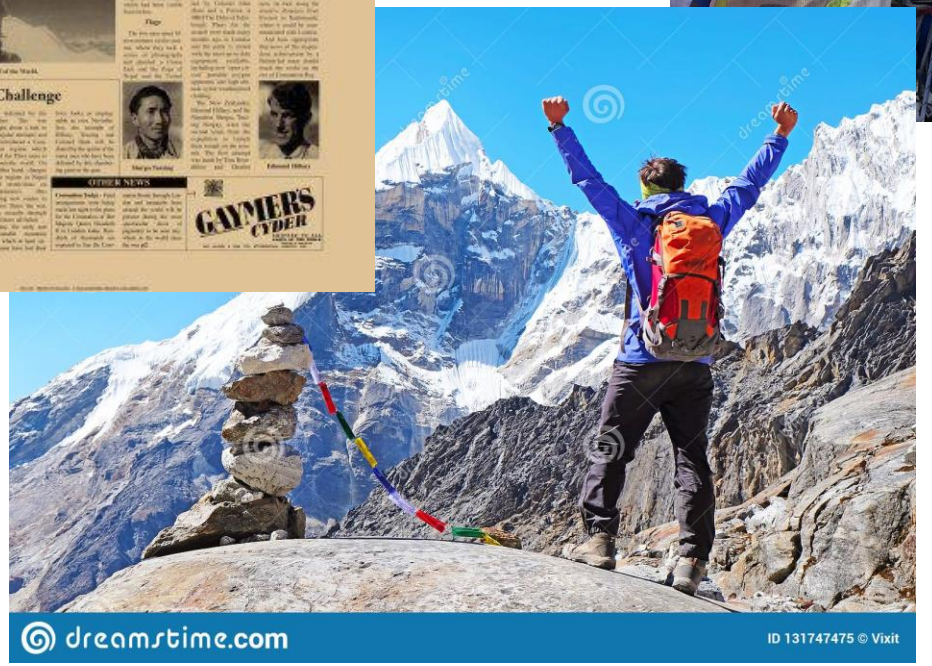
Contents

- Problem Statement
- Data
- Story
- Statistical Insights
- Machine Learning
- Findings

Problem Statement

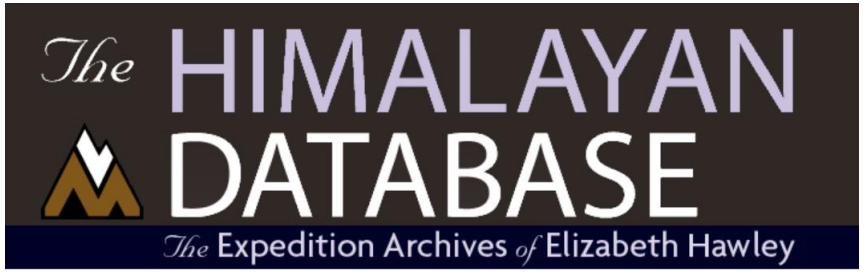
Predicting Himalayan Summit Success





Predicting Himalayan Summit Success

Data



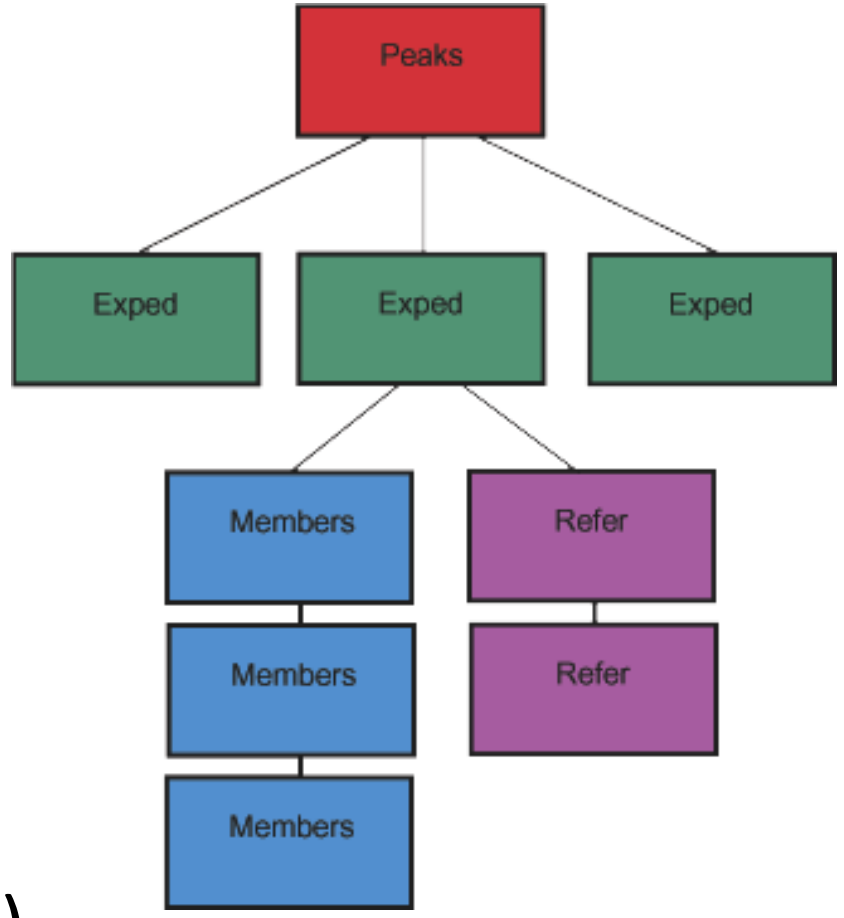
- Continued by Richard Salisbury
- 1905 to 2018
- Published bi-annually
- https://www.himalayandatabase.com/

Predicting Himalayan Summit Success

Data

The HIMALAYAN DATABASE
The Expedition Archives of Elizabeth Hawley

- Peaks – 450 (x 22)
- Expeditions – 10,000 (x 22)
- Members – 65,000 (x 85)
- Literature references (not used)
- Final merged dataset: (63,113 x 178)



Predicting Himalayan Summit Success

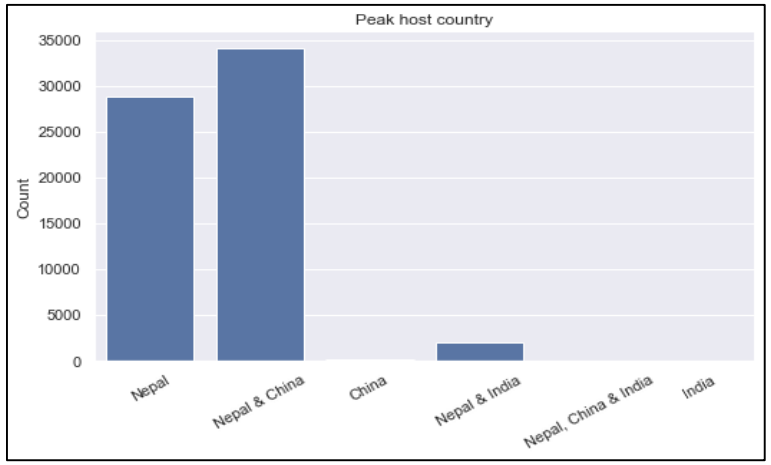
Data

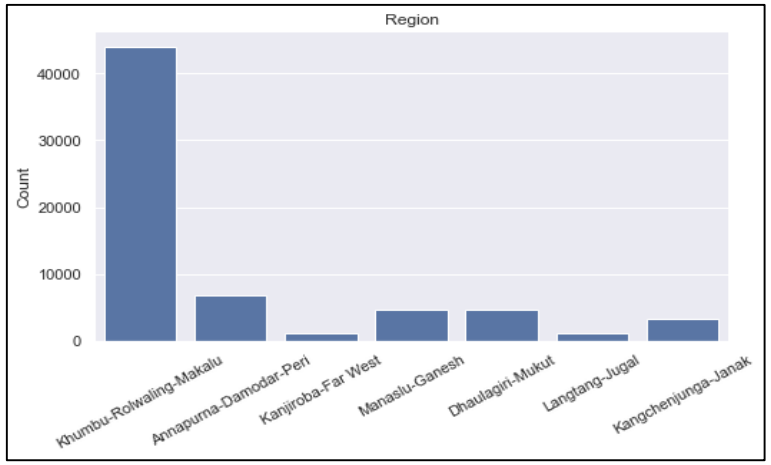
- Labels: Member success (Y/N)
- Drop columns
  - Multi-class (‘termreason’)
  - Non-categorical text
- Drop NaN rows < 1% of data
- Convert
  - Binary
  - Cyclical
  - Dummy variables

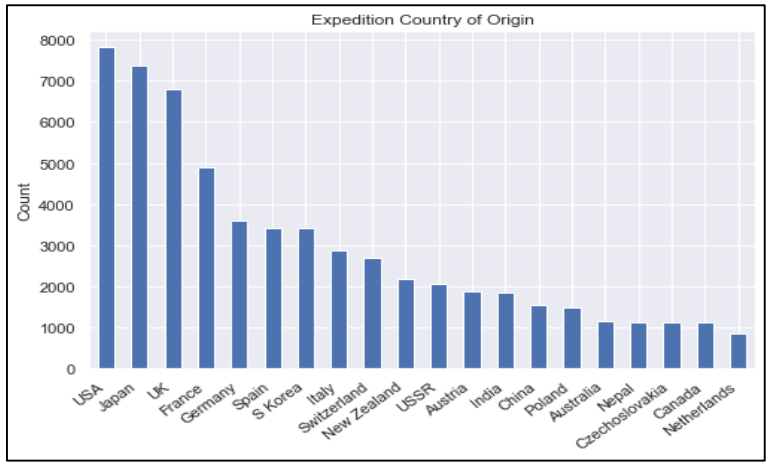
Features Data Type	63,113 x 51 (178)
Type	**Count**
Labelled data (binary)	1
Float (cyclical)	2
Categorical	7 (to 134 Dummy var)
Int	8
Binary	33

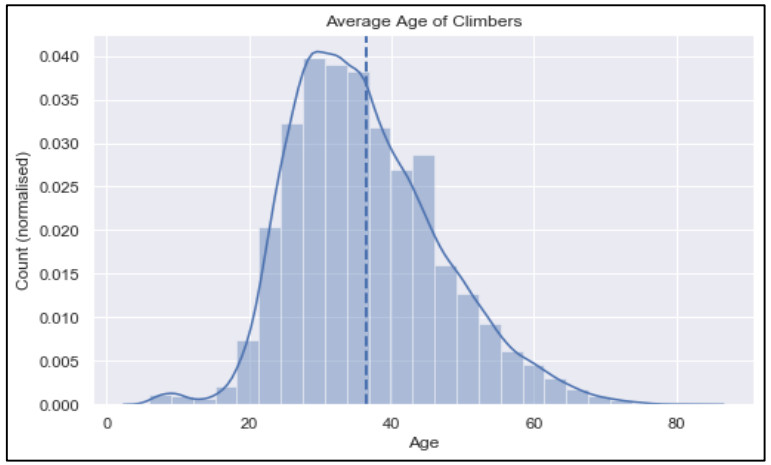
Story

Predicting Himalayan Summit Success



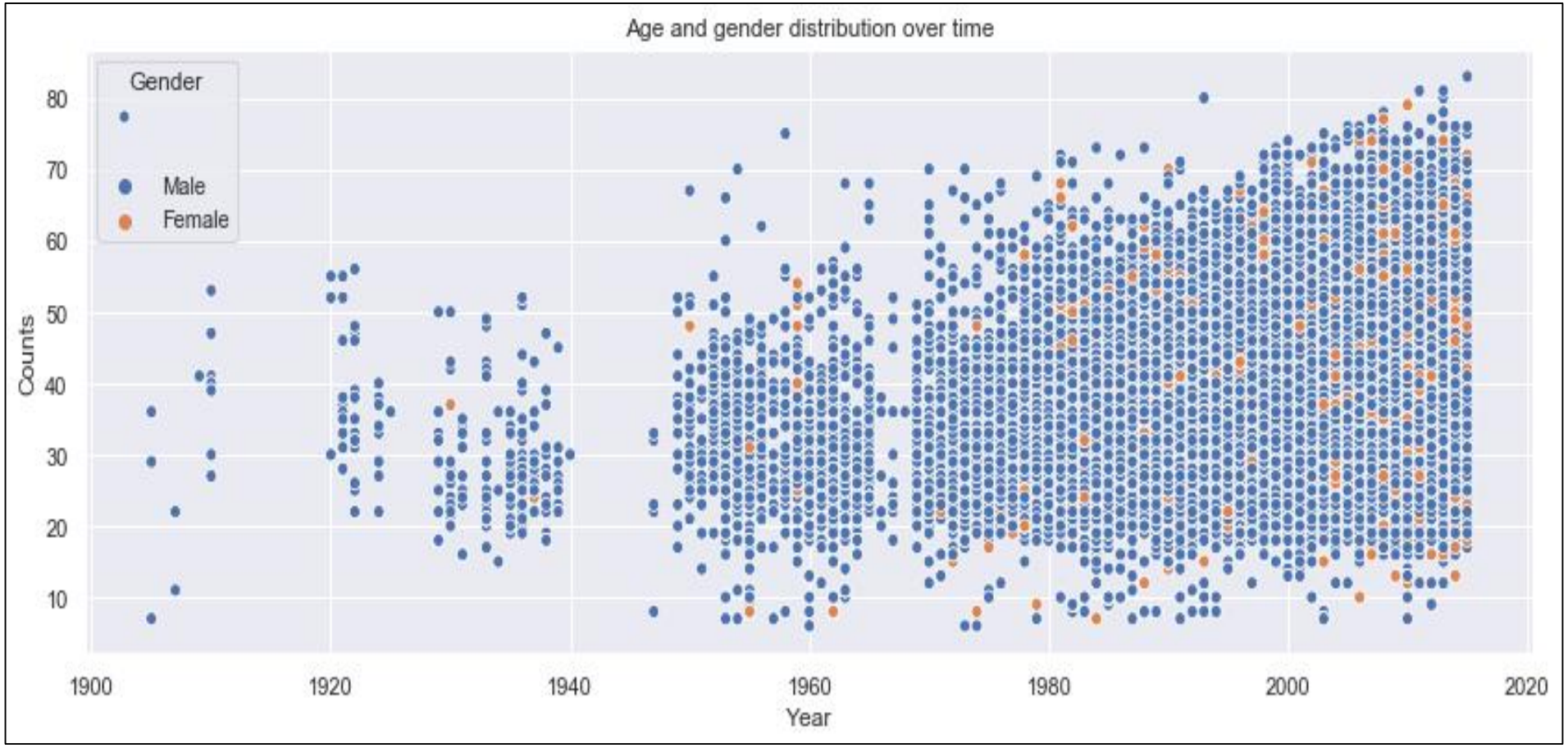






Story

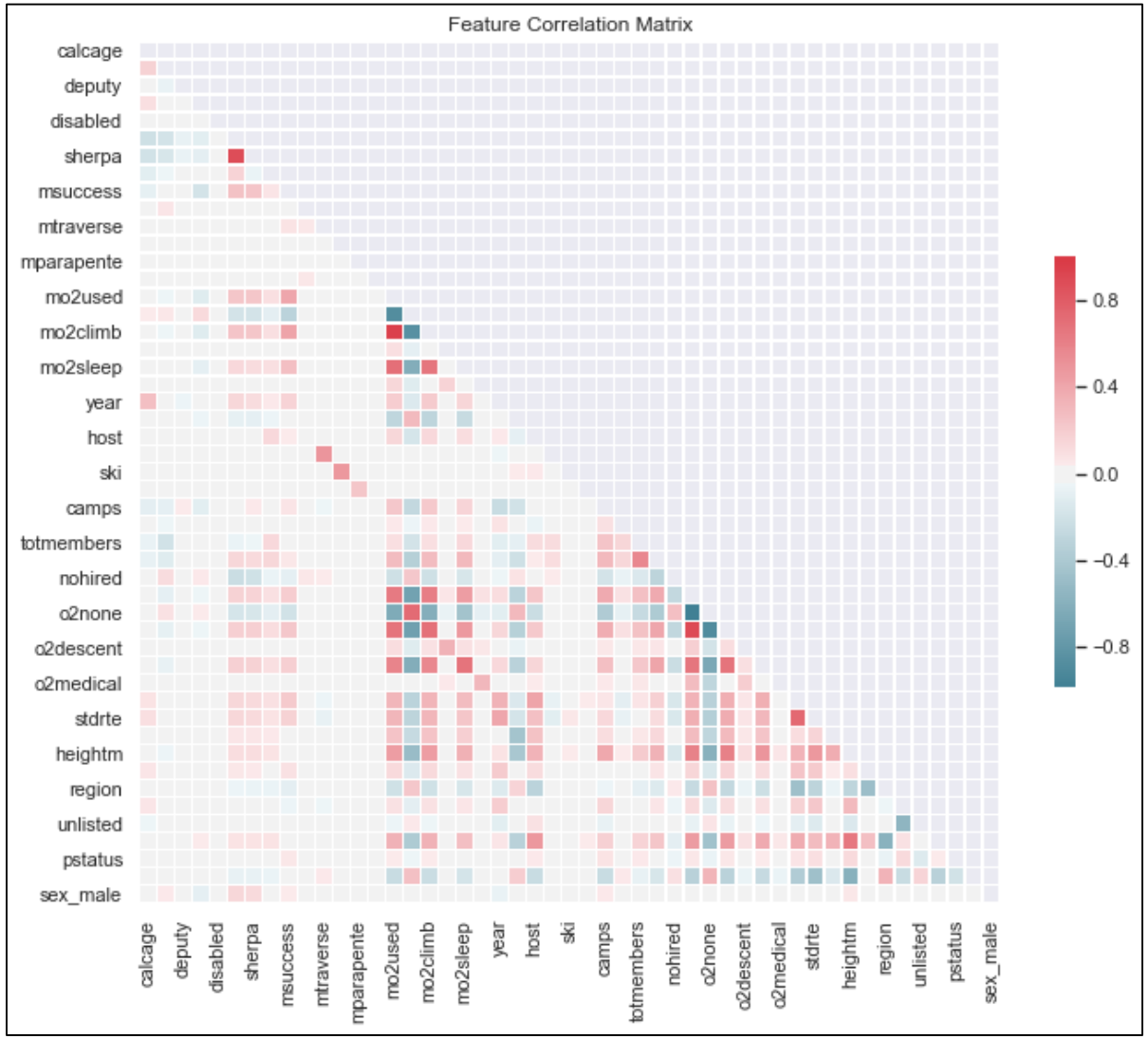
Predicting Himalayan Summit Success



Statistical Insights

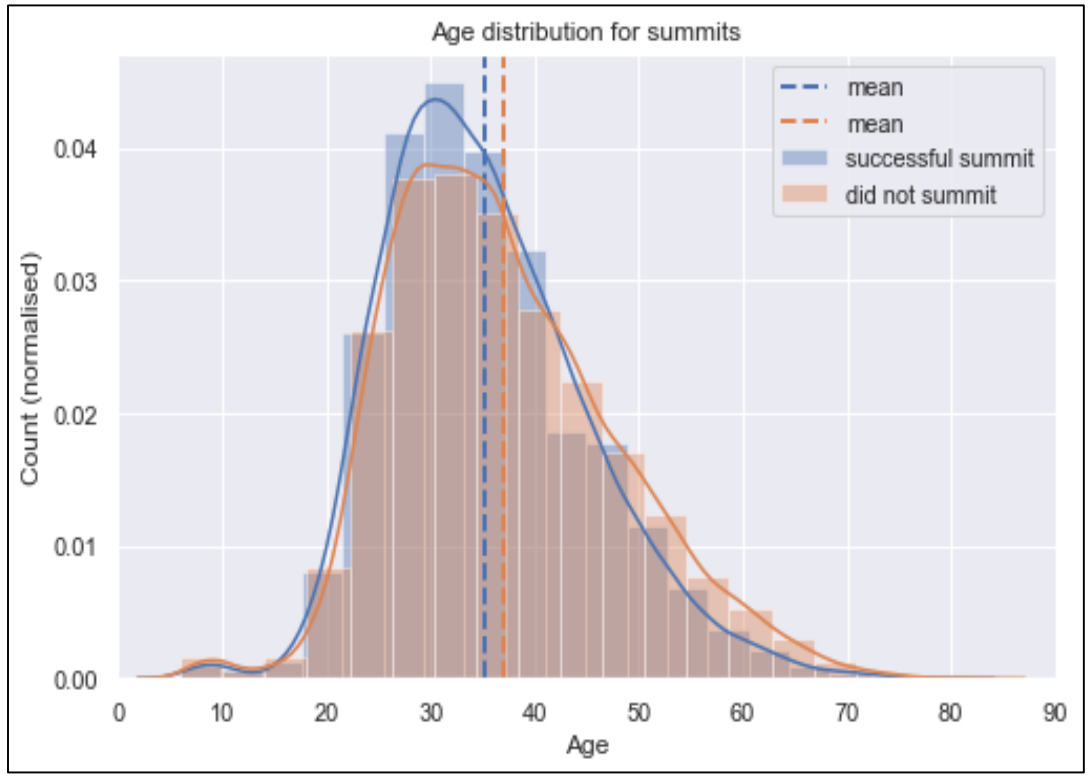
Predicting Himalayan Summit Success

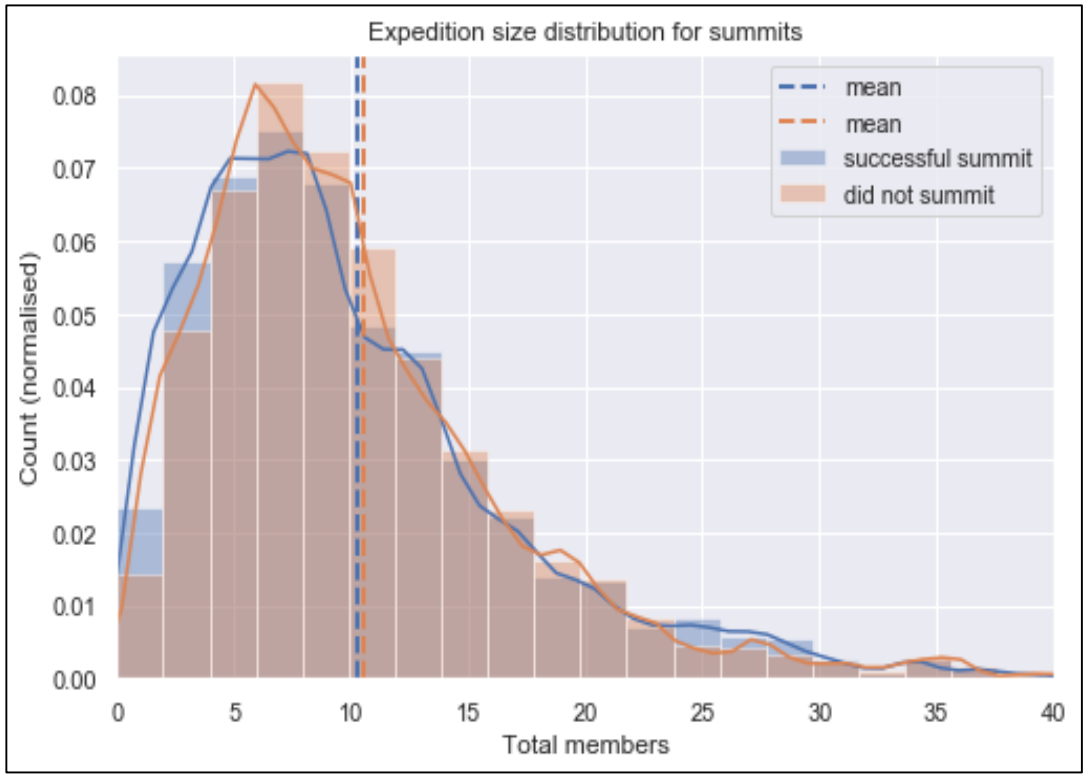
- msuccess
  - sherpa
  - mo2used/climb/sleep
  - year
  - stdrte



Statistical Insights

Predicting Himalayan Summit Success



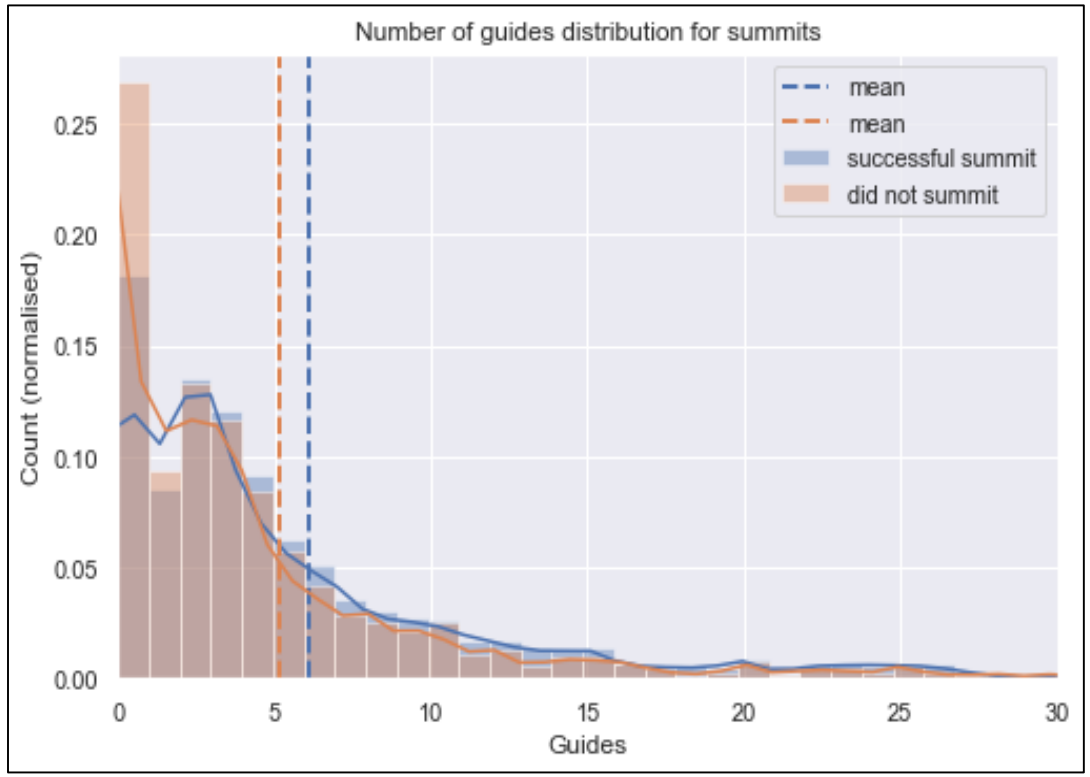


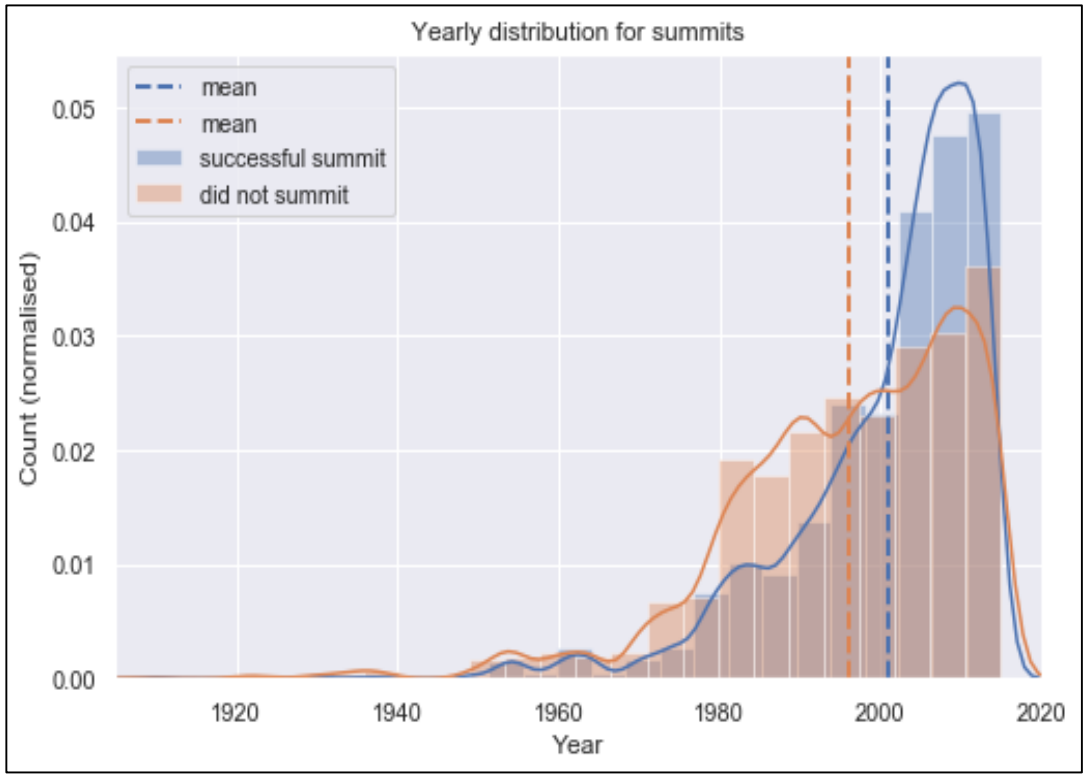
Age
μ = 35.26 / 37.02
p = 5.64e-93

Expedition size
μ = 10.23 / 10.52
p = 6.63e-05

Statistical Insights

Predicting Himalayan Summit Success





Number of guides
μ = 6.09 / 5.14
p = 7.30e-36

Year of expedition
μ = 1996 / 2001
p = 0.0

Machine Learning

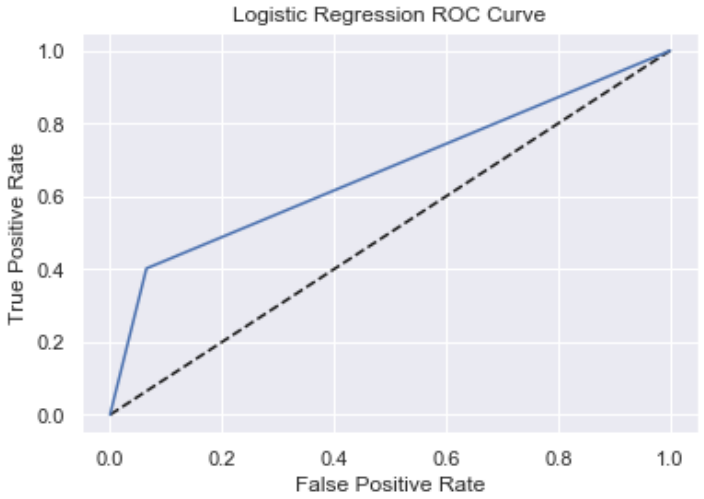
Predicting Himalayan Summit Success

- Heuristic
  - ‘comrte’ ~ 0.61
  - ‘mo2climb’ ~ 0.75
- KNN, LogReg, SVM
- Random Forest (1st)
- Gradient Boosting

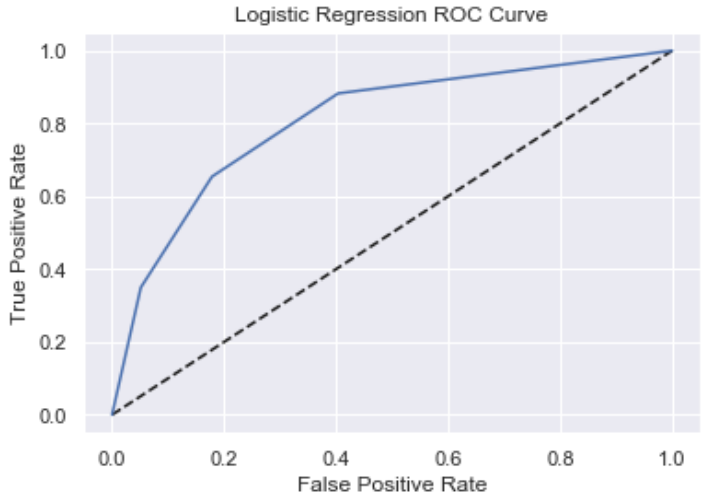
Classifier					
No.	**Estimator**	**Detail**	**Accuracy**	**ROC**	**Best parameters**
1	**GradientBoostingClassifier with gridsearch**	**GridSearchCV: n_estimators=[10,50,100] learning_rate=[0.01,0.1,0.5,1] max_depth=[10,100]**	**0.8454**	**0.9188**	**n_estimators=100 learning_rate=0.1 max_depth=10**
2	GradientBoostingClassifier	n_estimators=50 learning_rate=0.5 max_depth=10	0.8427	0.9139	
3	RandomForestClassifier	n_estimators=100	0.8379	0.9089	
4	ExtraTreesClassifier	n_estimators=100	0.8310	0.8804	
5	SVM with gridsearch for scaling and PCA	C=[1,10,100] gamma=[0.1,0.01] n_components=0.95	0.8178		C=100 gamma=0.01 n_components=.95
6	KNeighborsClassifier with scaling	StandardScaler n_neighbours=3	0.7938	0.8439	
7	AdaBoostClassifier	n_estimators=100	0.7895	0.8599	
8	KNeighborsClassifier with scaling and PCA	StandardScaler PCA (n_components=0.8) n_neighbours=3	0.7887	0.8371	
9	LogisticRegression with gridsearch for scaling and PCA	GridSearchCV: C=log[-5:8:5] penalty=['l1','l2'] n_components=[1,.95,.9,.85]	0.7743		C=100e6 penalty='l2' n_components=.95
10	KNeighborsClassifier	n_neighbours=3	0.7625	0.8071	
11	Heuristic	mo2climb' feature	0.7473	0.6684	

Machine Learning

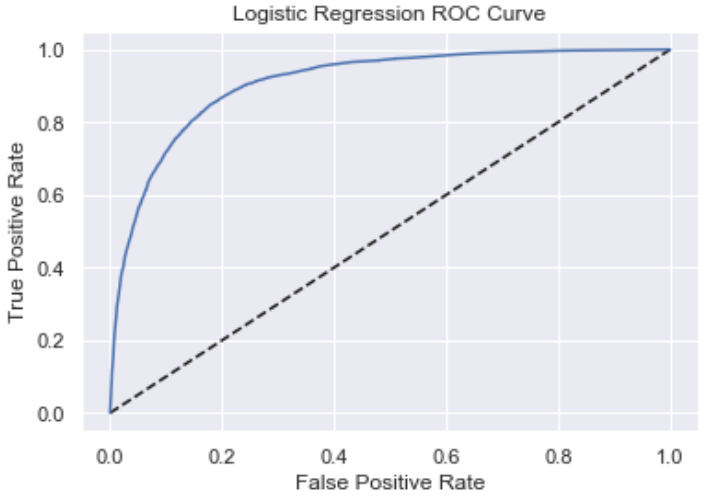
Predicting Himalayan Summit Success



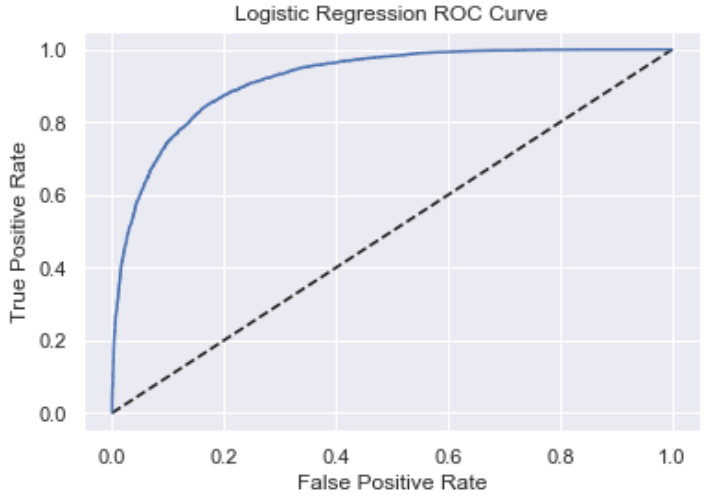
Heuristic



KNN



Random Forest

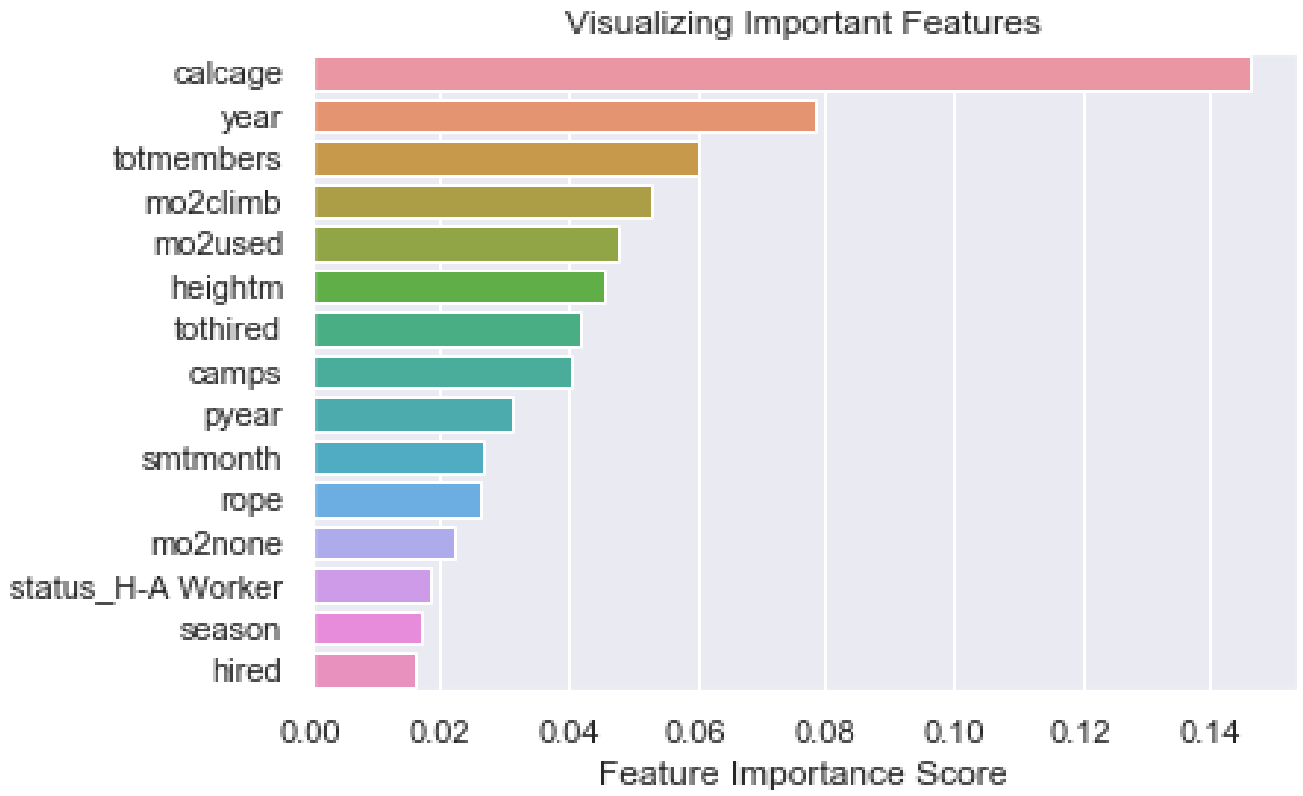


Gradient Boost

Findings

Predicting Himalayan Summit Success

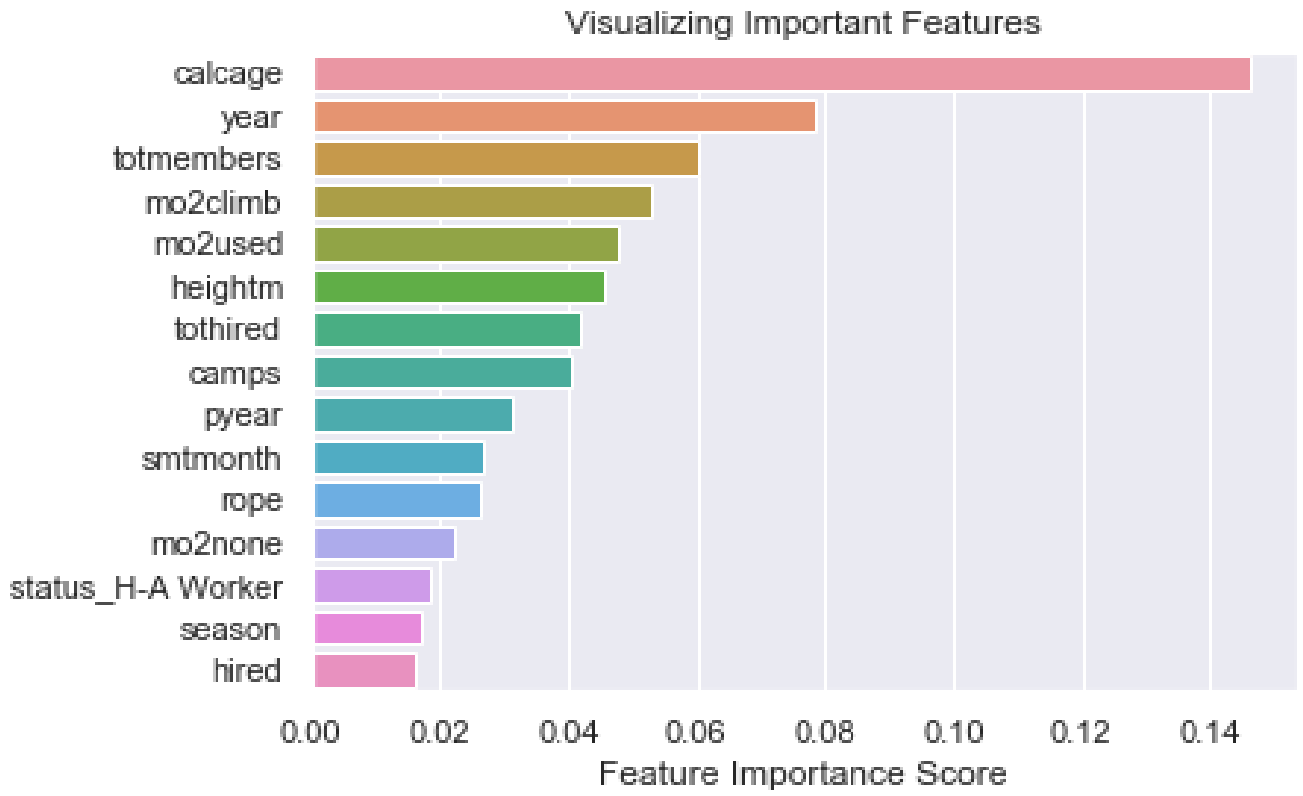
- Stats - Success
  - Age (younger)
  - Year (N/A, but later)
  - Expedition size (smaller)
- Heuristic ‘mo2climb’
- Guides (7th) (more)
- Look at next:
  - mo2used, heightm, camps, pyear, etc…



Findings

Predicting Himalayan Summit Success

- Recommendations:
  - Explore other features
  - Relative features
  - Supervised Learning
    - Cat. countries
    - Cat. occupations
    - Cat. status
  - Multi-class labels
    - ‘termreason’
    - Cat. for failures



Predicting Himalayan Summit Success

Thank you!

Jacques Poolman

Springboard