

PT or PT? Physical Therapy vs Personal Training

Jonathan Cheung
Project 3 (Subreddits)



Agenda

- 1) Overview: Physical Therapy and Personal Training
- 2) Work Flow - Data retrieval/cleaning, preprocessing, modeling
- 3) Modeling
- 4) Conclusion

Physical Therapy

- Physical therapists (PTs) are movement experts who optimize quality of life through prescribed exercise, hands-on care, and patient education.
- 3-year doctorate program



Personal Training

- An individual certified to have a varying degree of knowledge of general fitness involved in exercise prescription and instruction.
- Highschool Diploma / GED and a 3-month certification program

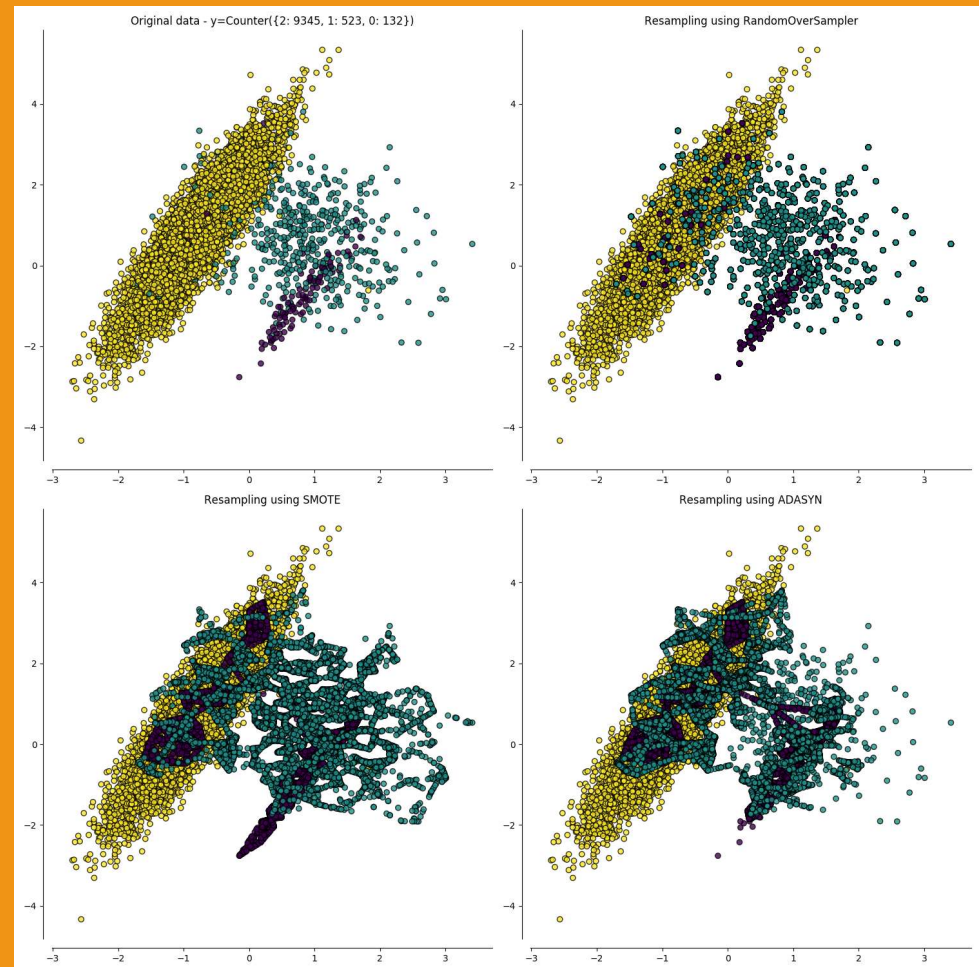


Workflow: Data retrieval, cleaning, preprocessing



Handling imbalanced classes

- 65% of data Class 1;
35% Class 0
- Random Oversampling
(my pick)
- SMOTE (Synthetic
Minority Oversampling
Technique)
- ADASYN (Adaptive
Synthetic Sampling)



Modeling

- LemmaTokenizer that filtered non-alphanumeric characters and English stop words
- Two main models:
 - CountVectorizer with Logistic Regression (L1 penalty)
 - N-gram range
 - Max features
 - CountVectorizer with Multinomial Naïve Bayes
 - N-gram range
 - Max features
 - Alpha (smoothing)

Results

Multinomial Naïve Bayes

Best Paramaters: {'cvec__max_features': 1500, 'cvec__ngram_range': (1, 2), 'multi_nb__alpha': 1e-10}

Best CV Score: 0.95

Raw Train Score: 0.96

Test Performance: 0.94

Sensitivity: 0.95

Specificity: 0.91

Precision: 0.95

Accuracy: 0.94

ROC AUC Score: 0.97

	Predicted Personal Training	Predicted Physical Therapy
Actual Personal Training	486	47
Actual Physical Therapy	50	937

Results

Logistic Regression

Best Paramaters: {'cvec__max_features': 1500, 'cvec__ngram_range': (1, 2)}

Best CV Score: 0.96

Raw Train Score: 0.99

Test Performance: 0.94

Sensitivity: 0.96

Specificity: 0.9

Precision: 0.95

Accuracy: 0.94

ROC AUC Score: 0.98

	Predicted Personal Training	Predicted Physical Therapy
Actual Personal Training	480	53
Actual Physical Therapy	42	945

Conclusion

Across several different models, these words held most influence:

Top 5 Physical Therapy Words: ['dpt', 'patient', 'physical', 'therapist', 'weak']

Top 5 Personal Training Words: ['client', 'fitness', 'nasm', 'trainer', 'training']