DATA 1030 Final Project: Anomaly Detection of Falling People

Kyle Lam

Department of Computer Science, Brown University

12/12/2024

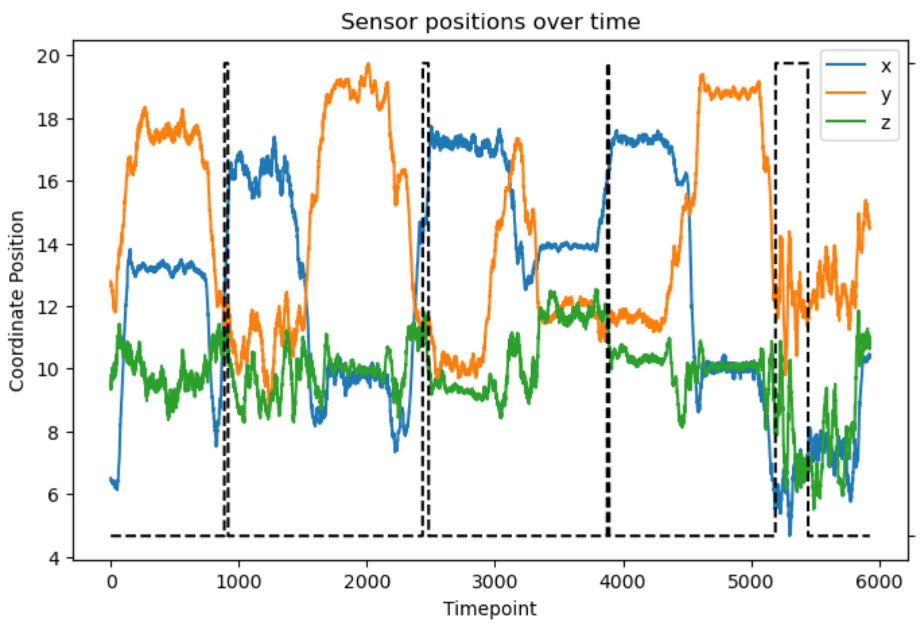
https://github.com/plumol/data1030-final-project

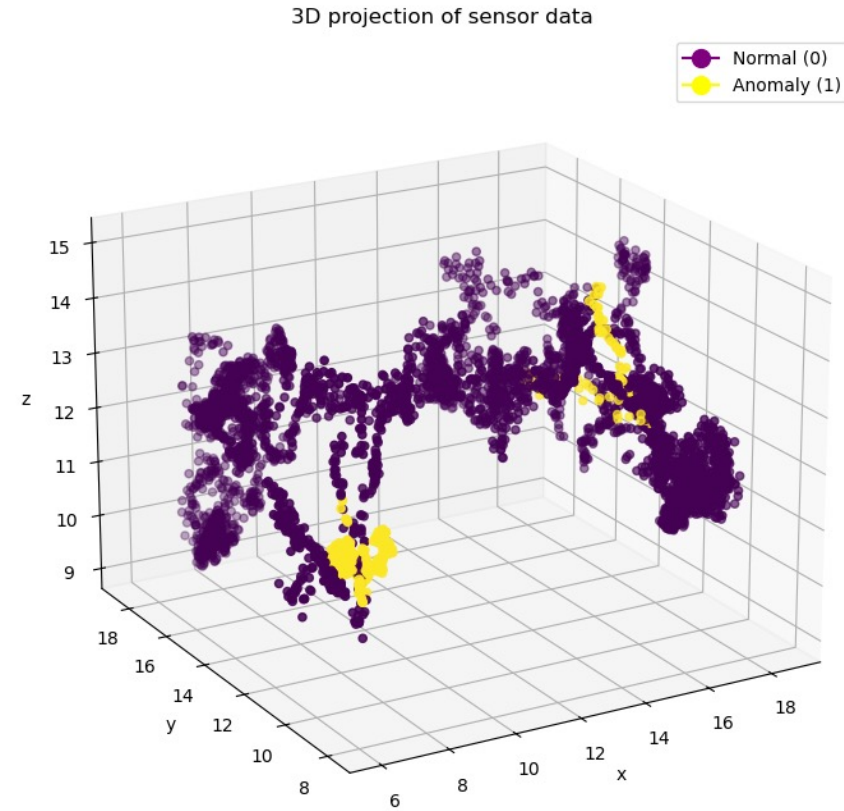
Dataset: **Localization Data for Person Activity**

- Anomaly detection: how can we predict for a very small positive class in a very unbalanced dataset?
- Background: This dataset[1] was created to develop strategies for safer "smart" care environments for the elderly.
  - Contains positional data from 4 physical sensors (ankles, belt, and chest area) performing various physical tasks.
  - Labels for "falling" are considered positive class while all other activities (walking, standing, sitting, laying down) are negative
- Problem: Classification!

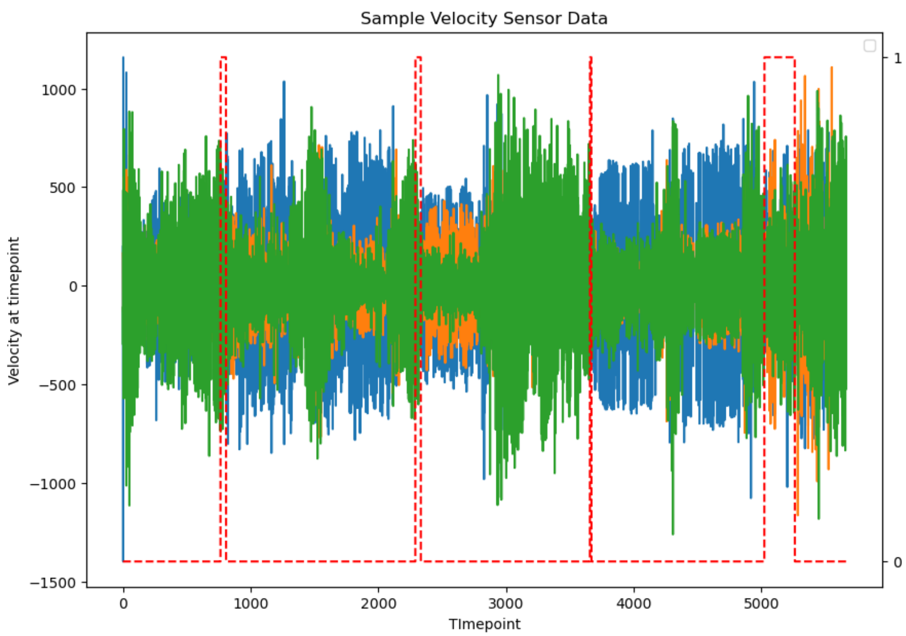
[1] Vidulin, V., Lustrek, M., Kaluza, B., Piltaver, R., & Krivec, J. (2010). Localization Data for Person Activity [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C57G8X.

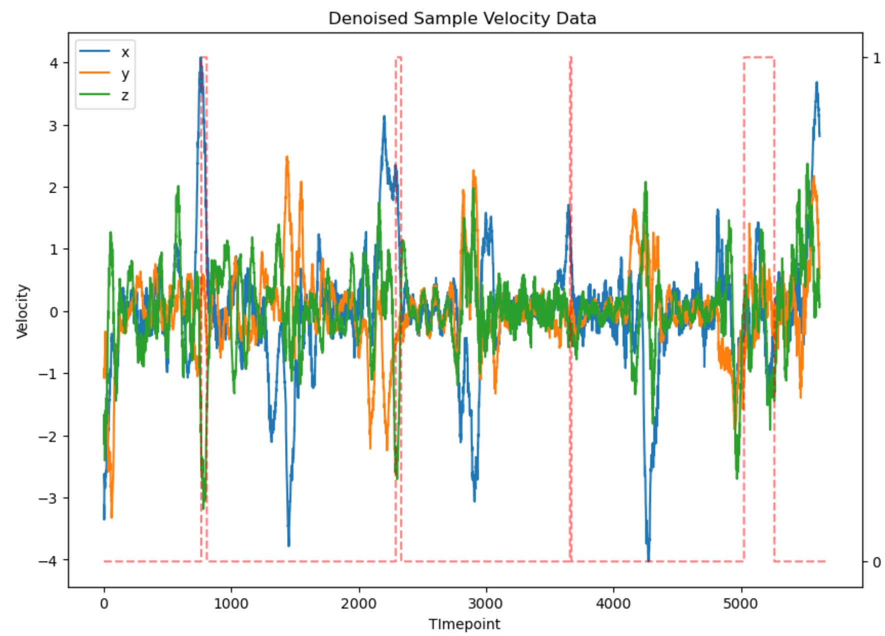
Recap EDA- Sensor Positions (2D + 3D)





Recap EDA- Sensor Velocity and Dataset





$$V = \frac{x_2 - x_1}{dt} \qquad dt = 0.027s$$

$$MA = \frac{1}{k} \sum_{\{i=n-k+1\}}^{n} p_i$$

k = window_size
p_i = value at i

Cross Validation and Model Training

- Dataset is split into 25 total users: 20 training/ 5 testing + shuffle for CV and 0.8 train-test-split

Type	Users	Samples	Normal %	Anomaly %
Training	20	134229	95.1% (127656)	4.9% (6573)
Testing	5	30030	94.6% (28420)	5.4% (1610)

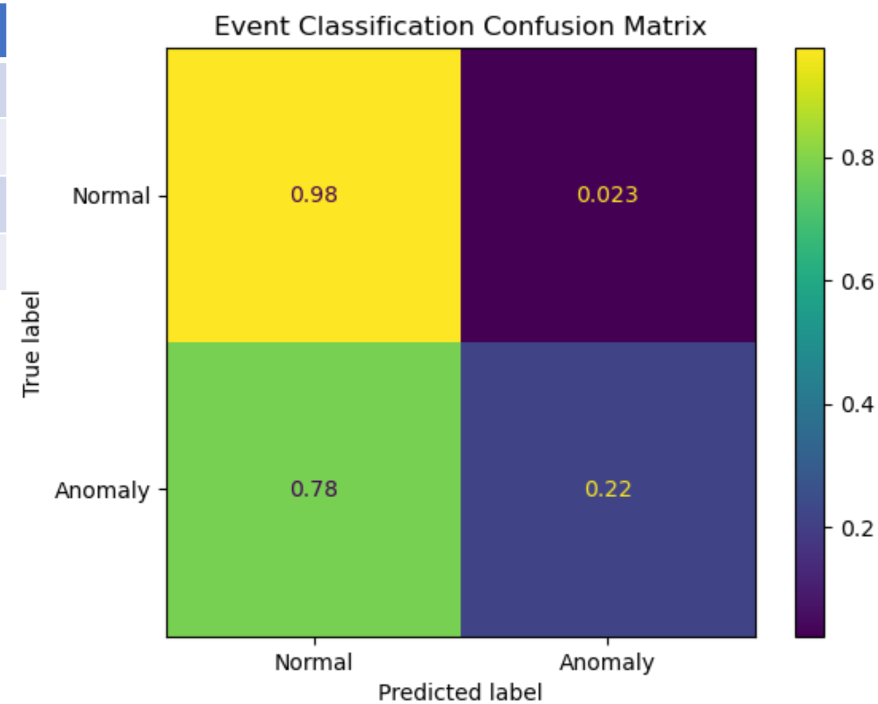
Algorithm	Parameters
LogisticRegression	C = [0.1, 1, 10]
RandomForest	max_features = [0.25, 0.5, 1.0] Max_depth = [5, 10, 30]
KNearestClassifier	n_neighbors = [1, 5, 10, 30] weights = [‘uniform’, ‘distance’]
XGBoostClassifier	n_estimators = [100, 500, 1000, 2000] max_depth = [3, 4, 5]

Results

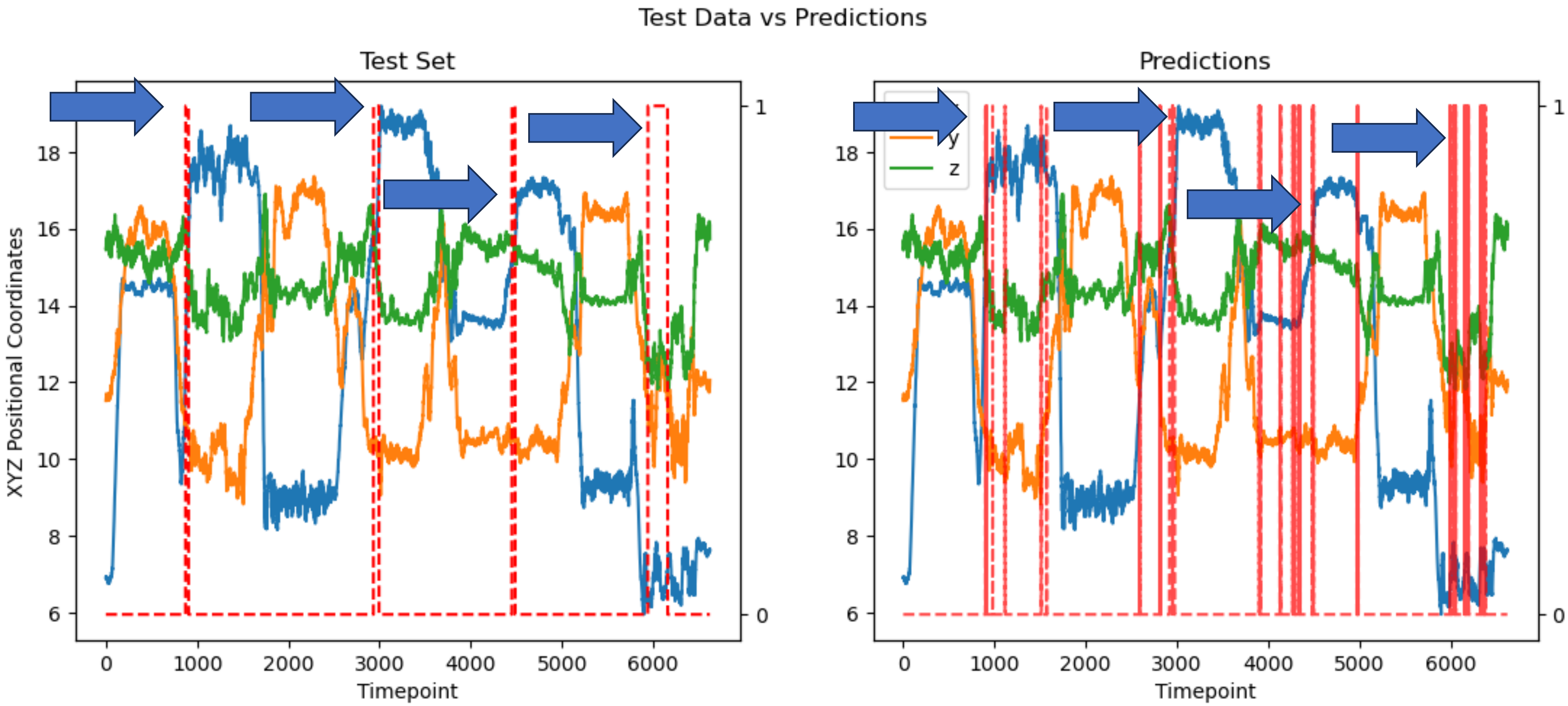
Model	F1 score mean	F1 score std
LogisticRegression	<0.0000	<0.0000
RandomForest	0.2562	0.0154
KNearestClassifier	0.2011	0.0112
XGBoostClassifier	**0.2857**	**0.0272**

Model	Best Parameters
RandomForest	max_depth = 30 max_features = 1.0
KNN	n_neighbors = 5 weights = ‘distance’

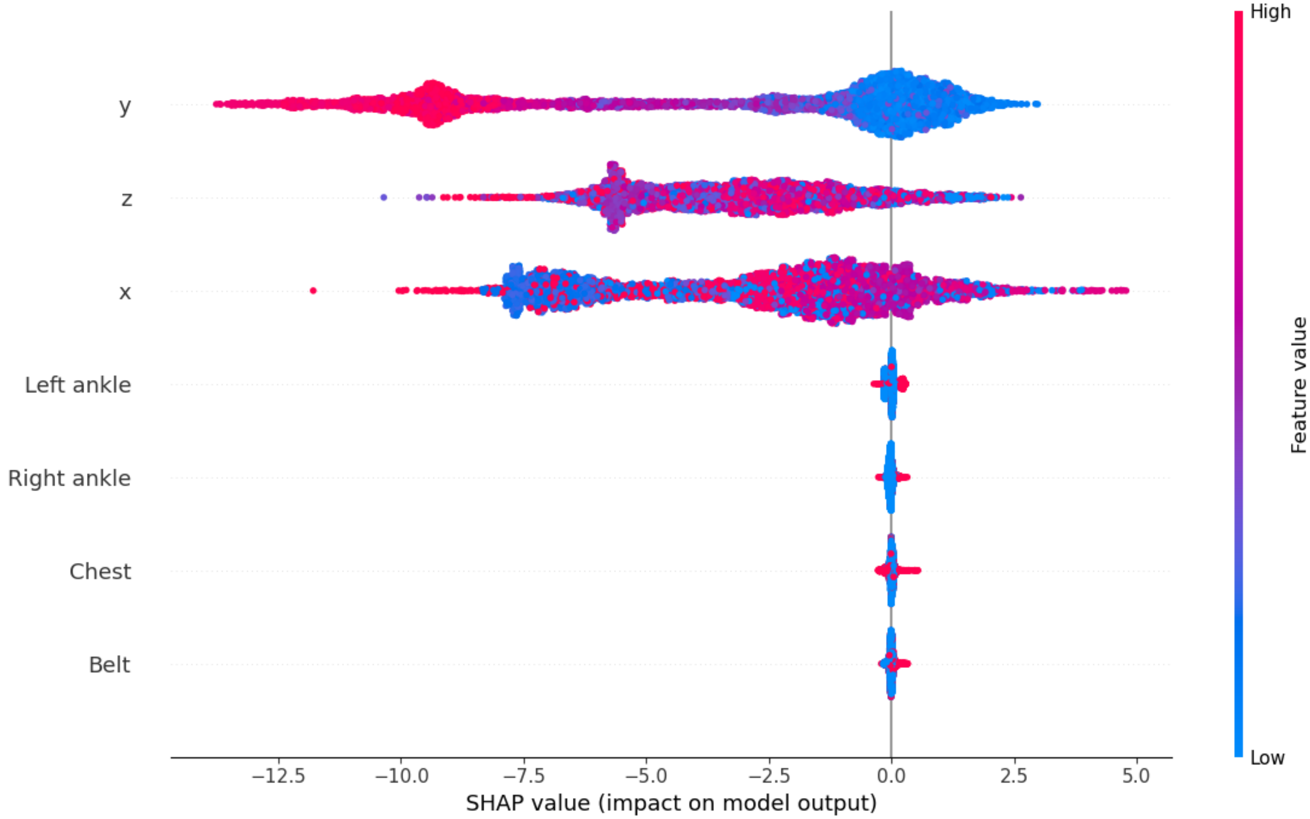
Model	Best Parameters
XGBoostClassifier	n_estimators = 2000 max_depth = 5



Predictions on a test dataset



Interpretability: SHAP Values



Future Directions

- Alternative training: velocity data or acceleration data may be better predictors for ‘falling’ events
  - May produce more meaningful interpretable results on SHAP
- More complex models: deep neural networks and specialized architectures (ex. Deep Autoencoding Gaussian Mixture Model)