

HUMAN ACTIVITIES PREDICTION

USING DATA MINING TECHNIQUE

CONTEXT

1 PROJECT OVERVIEW

2 PROBLEM SETTING & DEFINITION

- DATA SOURCE & DESCRIPTION
- DATA EXPLORATION

3. DATA MINING TASKS

- DATA CLEANING & PROCESSING
- DIMENSION REDUCTION

4. MODEL PERFORMANCE EVALUATION & SELECTION

5. PROJECT RESULTS & INSIGHTS

- PERFORMANCE EVALUATION RESULTS
- IMPACT OF PROJECT OUTCOMES
- CONCLUSIONS & FUTURE WORK

PROBLEM SETTING

- The problem is situated in the context of recognition of daily & sports activity.
- Understanding the significant deviation in signals on different activity
- The goal is to understand and predict the activity based on signals given by the sensors.

DATA SOURCE & DESCRIPTION

SOURCE: UC Irvine Machine Learning Repository

Citation: Billur Barshan and Kerem Atun

https://archive.ics.uci.edu/dataset/256/daily+and+sports+activities

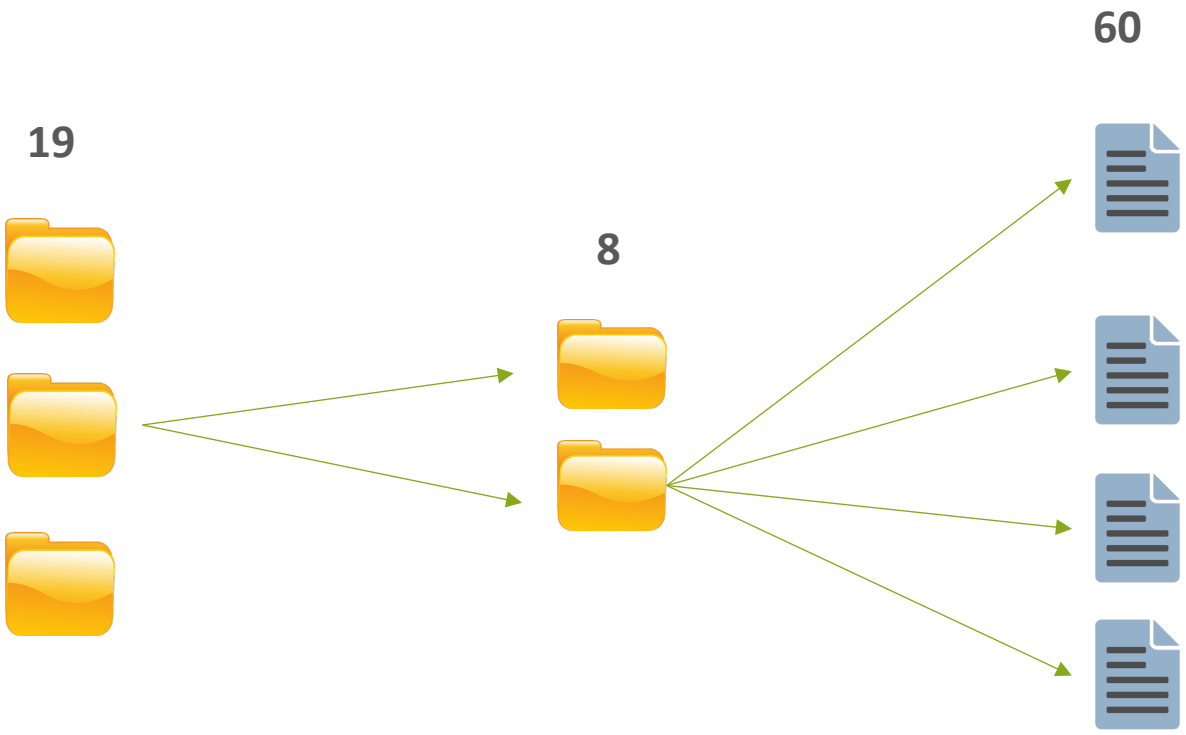
Target variable: multinomial variable indicating which activity is performed by the person

Predictors: 45 numeric variables

Dataset shape: 1140000 rows and 47 columns

- Each of the 19 activities is performed by eight subjects.
- subjects (4 female, 4 male, between the ages 20 and 30)
- Total signal duration is 5 minutes for each activity of each subject.

DATA SOURCE & DESCRIPTION

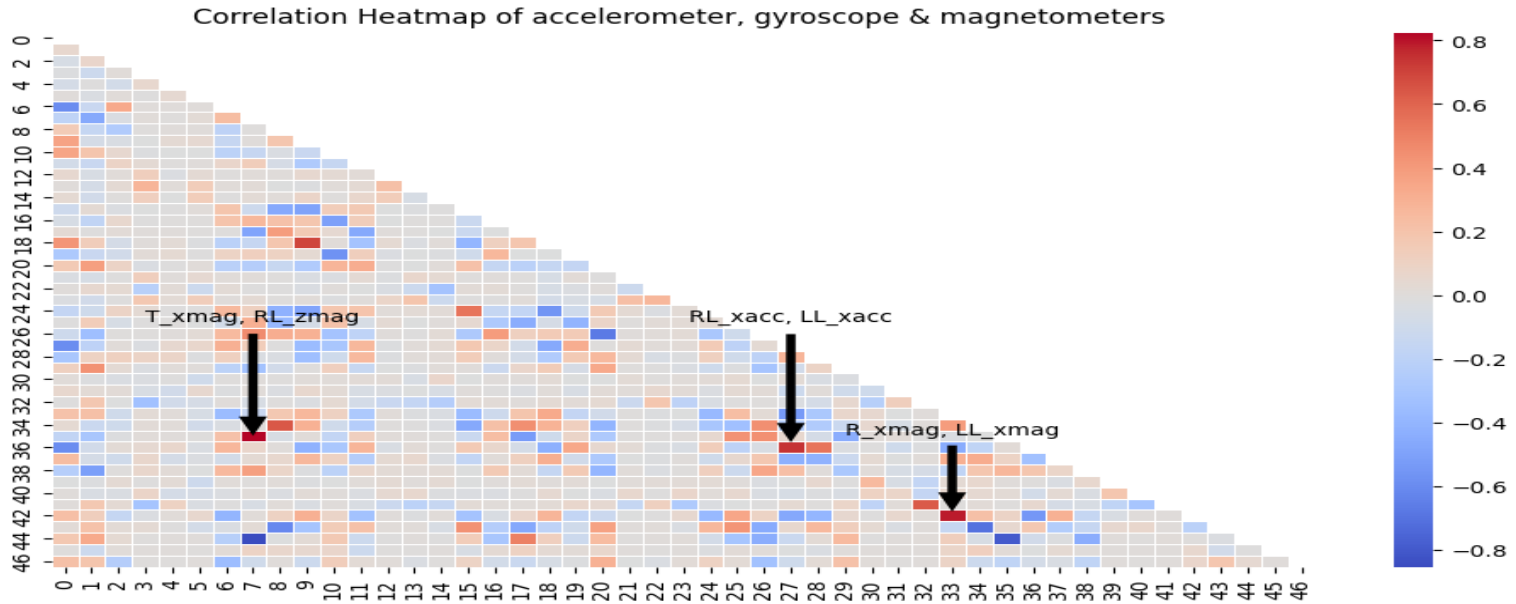


DATA COLLECTION

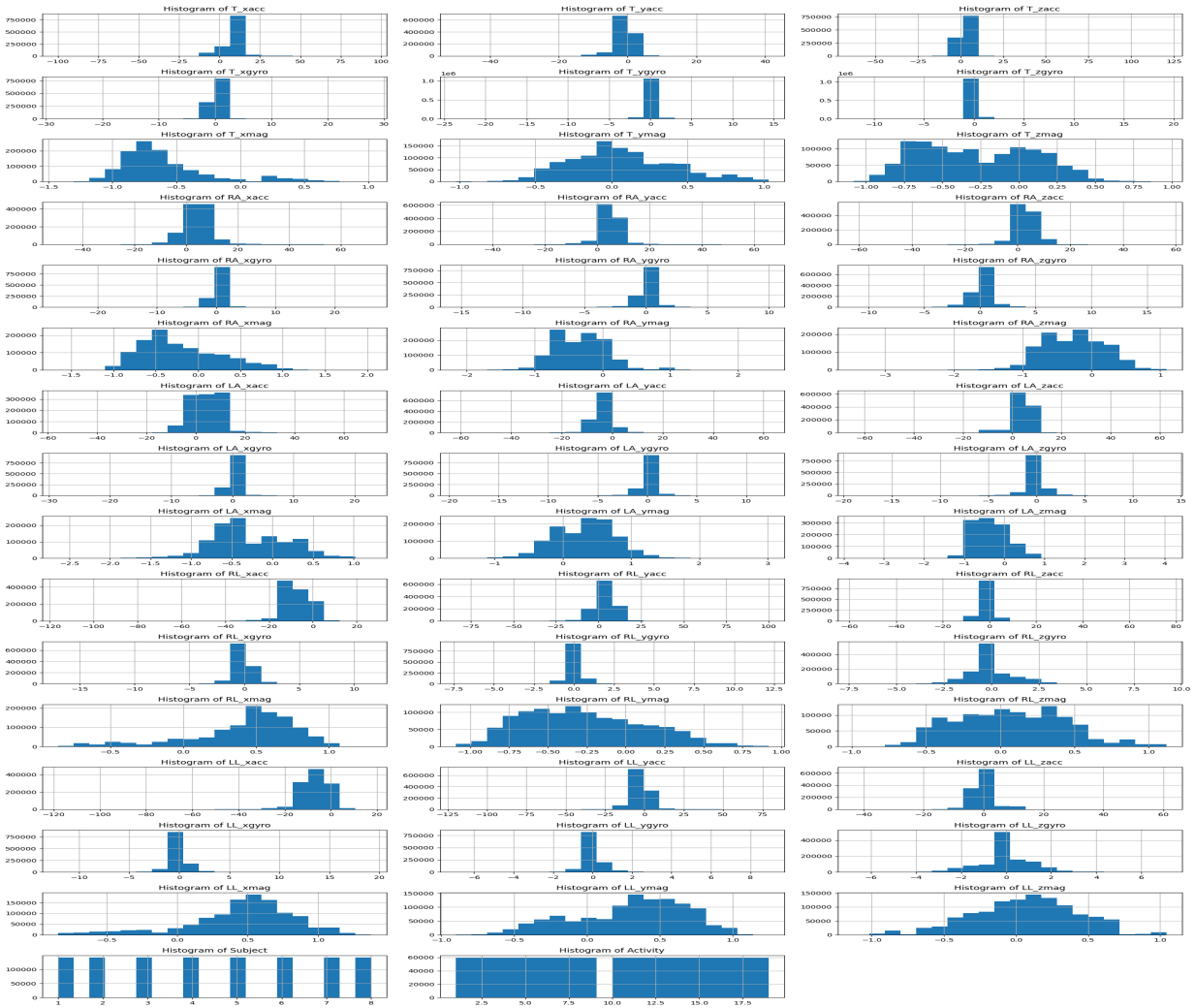
EXPLORATORY DATA ANALYSIS

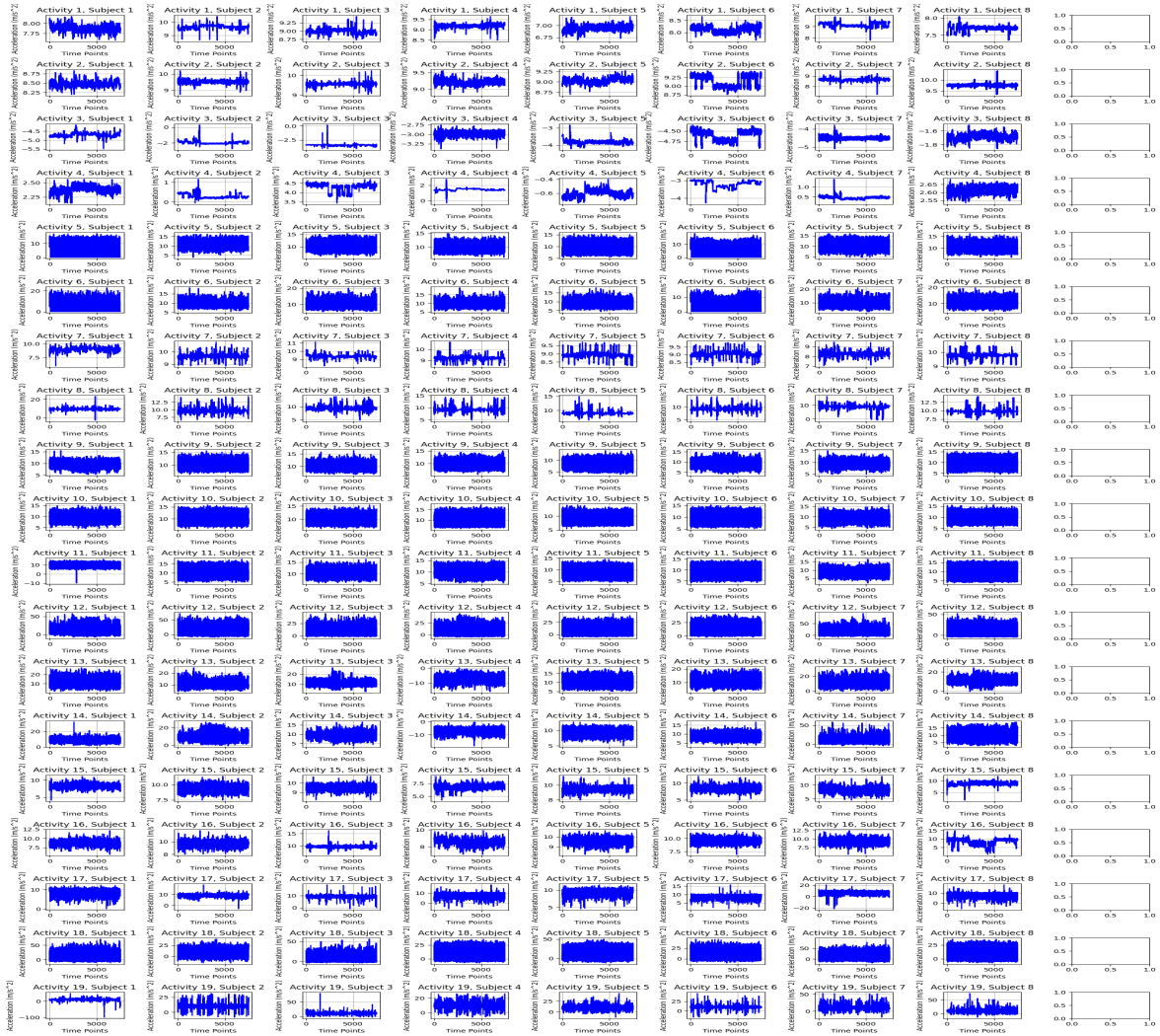
shape: (9, 48)

describe	T_xacc	T_yacc	T_zacc	T_xgyro	T_ygyro	T_zgyro	T_xmag	T_ymag	T_zmag	RA_xacc	RA_yacc	RA_zacc	RA_xgyro	RA_ygyro	RA_zgyro	RA_xmag
str	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64
"count"	1.14e6	1.14e6	1.14e6	1.14e6	1.14e6	1.14e6	1.14e6	1.14e6	1.14e6	1.14e6	1.14e6	1.14e6	1.14e6	1.14e6	1.14e6	1.14e6
"null_count"	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
"mean"	7.765766	-0.811036	2.768845	-0.002796	0.013695	-0.003312	-0.598436	0.061729	-0.272517	4.260595	4.938779	3.119262	0.016302	-0.021484	-0.00318	-0.21123
"std"	5.637887	2.623027	3.53826	0.794011	0.69104	0.310766	0.3561	0.340436	0.373412	5.821339	4.580221	3.869474	0.869443	0.764749	1.023147	0.46358
"min"	-99.715	-49.941	-62.664	-27.851	-23.598	-12.067	-1.4226	-1.0228	-1.0806	-49.535	-53.915	-62.145	-26.663	-15.439	-11.35	-1.6513
"25%"	6.907	-1.5095	0.89985	-0.16289	-0.10248	-0.09748	-0.8247	-0.17883	-0.60172	0.094976	2.5398	0.81226	-0.1356	-0.16637	-0.2033	-0.54955
"50%"	8.8303	-0.38953	2.7037	0.000461	0.017438	-0.002664	-0.68975	0.0225	-0.30926	4.143	4.5996	2.6645	0.007994	-0.00421	-0.001273	-0.33126
"75%"	9.6905	0.41362	4.4052	0.16482	0.13131	0.088826	-0.49252	0.28235	0.040582	8.505	7.491	5.8893	0.16181	0.14566	0.18436	0.1036
"max"	93.694	41.013	120.53	27.671	14.379	19.262	1.0215	1.0309	0.96339	71.652	65.427	56.384	26.134	9.9733	16.734	2.0433



EXPLORATORY DATA ANALYSIS





MODEL PERFORMANCE

KNN

	precision	recall	f1-score	support
0	1.00	1.00	1.00	11913
1	1.00	1.00	1.00	11946
2	1.00	1.00	1.00	12037
3	1.00	1.00	1.00	12142
4	0.91	1.00	0.95	12006
5	0.97	0.93	0.95	11999
6	0.99	1.00	1.00	12299
7	0.96	0.97	0.96	11926
8	0.97	1.00	0.99	11960
9	0.97	0.99	0.98	11999
10	0.98	0.98	0.98	12078
11	0.97	0.99	0.98	11973
12	0.97	0.99	0.98	12081
13	0.98	1.00	0.99	11950
14	1.00	1.00	1.00	11941
15	1.00	1.00	1.00	11962
16	1.00	1.00	1.00	11889
17	0.95	0.94	0.95	12097
18	0.96	0.81	0.88	11802
accuracy			0.98	228000
macro avg	0.98	0.98	0.98	228000
weighted avg	0.98	0.98	0.98	228000

Decision Tree

Accuracy: 0.94

	precision	recall	f1-score	support
0	1.00	1.00	1.00	11913
1	1.00	1.00	1.00	11946
2	1.00	1.00	1.00	12037
3	1.00	1.00	1.00	12142
4	0.94	0.90	0.92	12006
5	0.88	0.94	0.91	11999
6	0.97	0.98	0.97	12299
7	0.87	0.88	0.87	11926
8	0.94	0.95	0.95	11960
9	0.91	0.91	0.91	11999
10	0.91	0.92	0.92	12078
11	0.94	0.93	0.94	11973
12	0.93	0.93	0.93	12081
13	0.95	0.96	0.96	11950
14	0.99	0.99	0.99	11941
15	0.99	0.99	0.99	11962
16	1.00	1.00	1.00	11889
17	0.88	0.87	0.87	12097
18	0.75	0.71	0.72	11802
accuracy			0.94	228000
macro avg	0.94	0.94	0.94	228000
weighted avg	0.94	0.94	0.94	228000

Random Forest

	precision	recall	f1-score	support
1	1.00	0.00	0.00	17850
2	0.00	0.00	0.00	17911
3	1.00	0.99	1.00	18019
4	1.00	1.00	1.00	17915
5	0.00	0.00	0.00	18010
6	0.00	0.00	0.00	17948
7	0.00	0.00	0.00	17989
8	0.75	0.00	0.00	18197
9	0.91	0.00	0.00	18064
10	0.29	0.83	0.43	17990
11	0.00	0.00	0.00	18028
12	0.73	0.08	0.14	18008
13	0.17	0.93	0.28	17861
14	0.90	0.00	0.01	17971
15	0.16	0.99	0.27	18171
16	0.82	0.95	0.88	18091
17	1.00	0.82	0.90	17990
18	0.73	0.08	0.14	17978
19	0.59	0.01	0.02	18009
accuracy			0.35	342000
macro avg	0.53	0.35	0.27	342000
weighted avg	0.53	0.35	0.27	342000

Without Transformation: 67%
Dropping Highly correlated features: 89%
Implementation PCA: 98%

CONCLUSION

- KNN demonstrated superior performance with 98% accuracy in classifying daily and sports activities using motion sensor data.
- Future efforts will focus on refining the KNN model through hyperparameter tuning and feature engineering to optimize its performance further.
- This research highlights the potential of machine learning in health monitoring, sports analytics, and human-computer interaction, showcasing practical applications for automated activity recognition with high accuracy and reliability.

THANK YOU