

HUMAN ACTIVITIES PREDICTION

USING DATA MINING TECHNIQUE

CONTENT

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.

PURPOSE

Health and Fitness Monitoring:

- Predicting activities using sensor data can be used in health and fitness applications to monitor and analyze physical activities of individuals.
- It can help in designing personalized fitness plans, tracking workout sessions, and ensuring proper exercise form.

Sports Performance Analysis:

- Sports scientists and coaches can use activity prediction models to analyze athletes' movements during training and competitions.
- This can aid in optimizing training programs, improving techniques, and preventing injuries.

DATA SOURCE

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: 1.14Mn rows and 47 columns

DATA DESCRIPTION

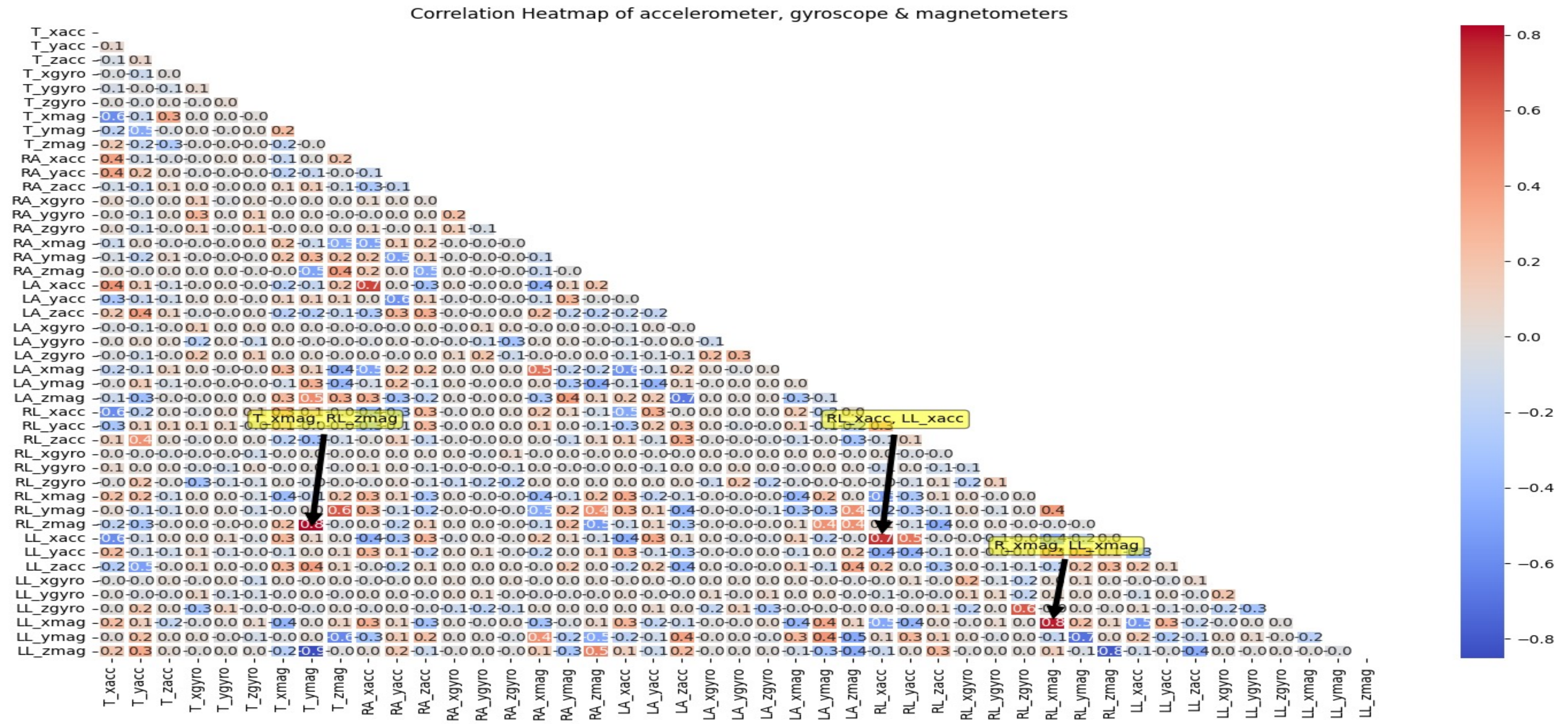
- Each of the 19 activities is performed by eight subjects.
- subjects (4 female, 4 male, between the ages 20 and 30)
- Total signal duration is 5 minutes for each activity of each subject.
- Sensors :

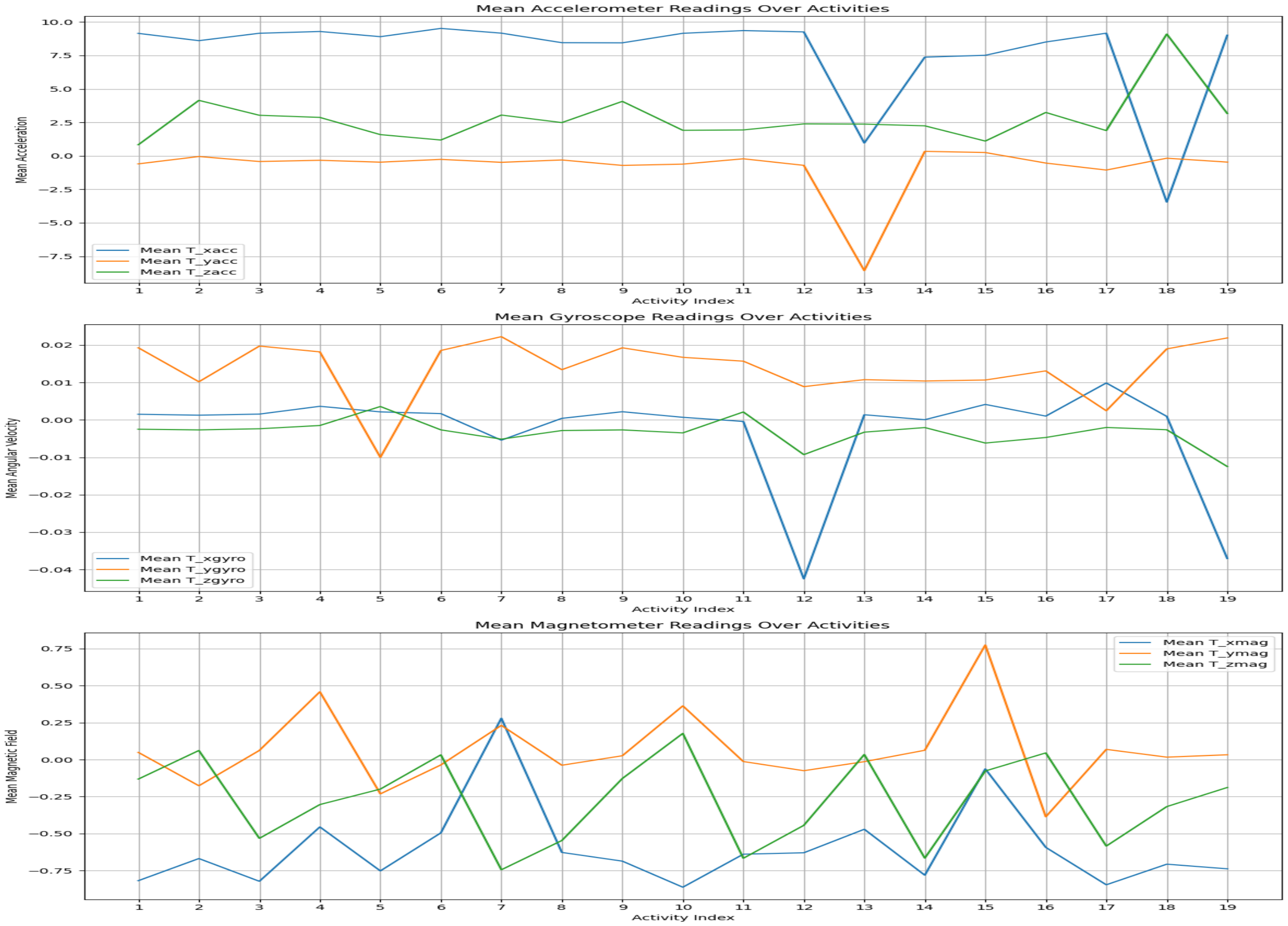
  - Accelerometer (T_xacc, T_yacc, T_zacc)
  - Gyroscope (T_xgyro, T_ygyro, T_zgyro)
  - Magnetometer (T_xmag, T_ymag, T_zmag)

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





DATA ANALYSIS

DATA TRANSFORMATION

DOWN-SAMPLELING

AGGREGATION

UFE

PRINCIPAL COMPONENT ANALYSIS (PCA)

CLUSTERING

T-DIS STOCHASTIC NEIGBOUR SAMPLEING

TIME-BASED SAMPLING

DATA TRANSFORMATION

AGGREGATION

Sampling Frequency in HZ

Activity Performed for 5min

Persons

Activity

25 x (5 x 60) x 8 x 19 = 11,40,000 records

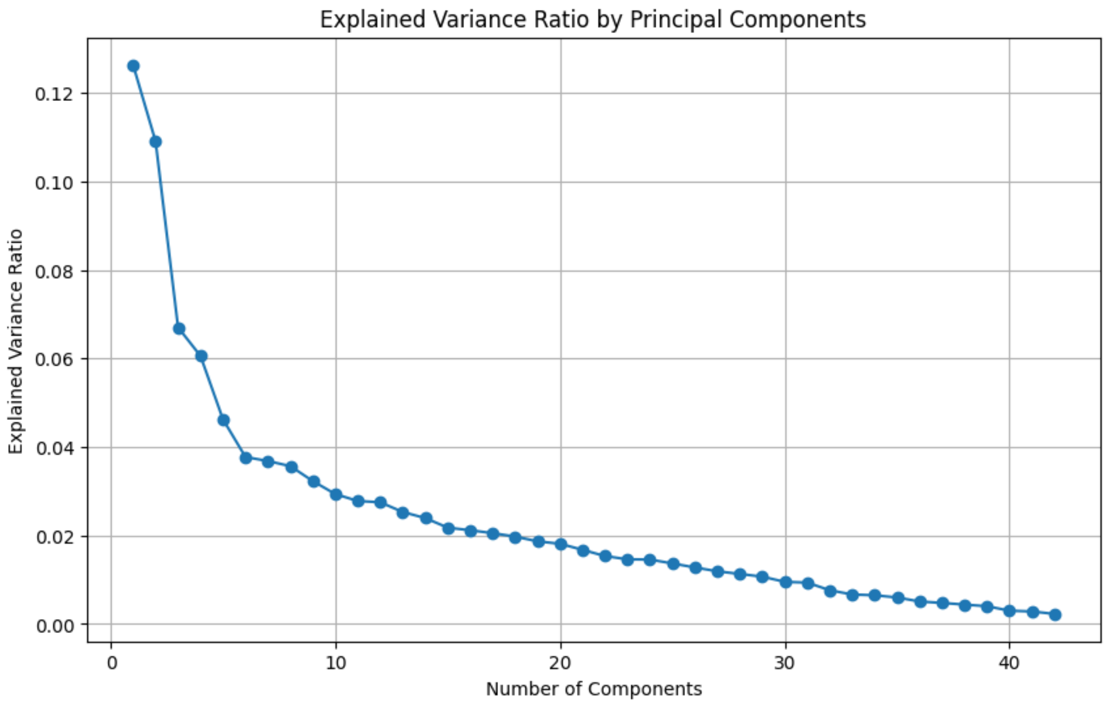
1 sec → **25 records**

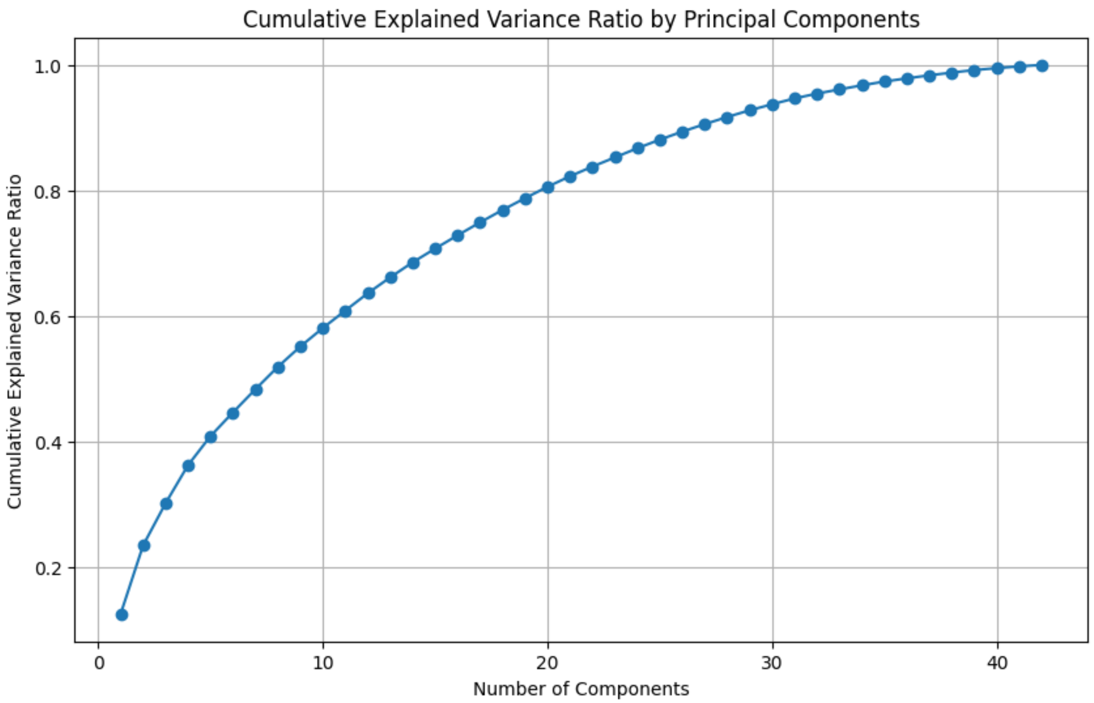
GROUP BY
- Mean
- Min
- Max
- Std

```
Number of Rows: 45600
Number of Columns: 182
```

FEATURE SELECTION

PRINCIPAL COMPONENT ANALYSIS (PCA)





MODEL PERFORMANCE

Decision Tree

Classification Report:

	precision	recall	f1-score	support
1	0.82	0.87	0.84	11913
2	0.44	0.41	0.43	11946
3	0.97	1.00	0.98	12037
4	1.00	1.00	1.00	12142
5	0.67	0.89	0.76	12006
6	0.60	0.85	0.70	11999
7	0.53	0.42	0.47	12299
8	0.50	0.41	0.45	11926
9	0.41	0.49	0.44	11960
10	0.60	0.63	0.61	11999
11	0.42	0.37	0.40	12078
12	0.59	0.61	0.60	11973
13	0.80	0.82	0.81	12081
14	0.77	0.93	0.84	11950
15	0.94	0.91	0.92	11941
16	0.96	0.96	0.96	11962
17	0.98	0.97	0.98	11889
18	0.40	0.22	0.28	12097
19	0.34	0.23	0.28	11802
accuracy			0.68	228000
macro avg	0.67	0.68	0.67	228000
weighted avg	0.67	0.68	0.67	228000

Random Forest

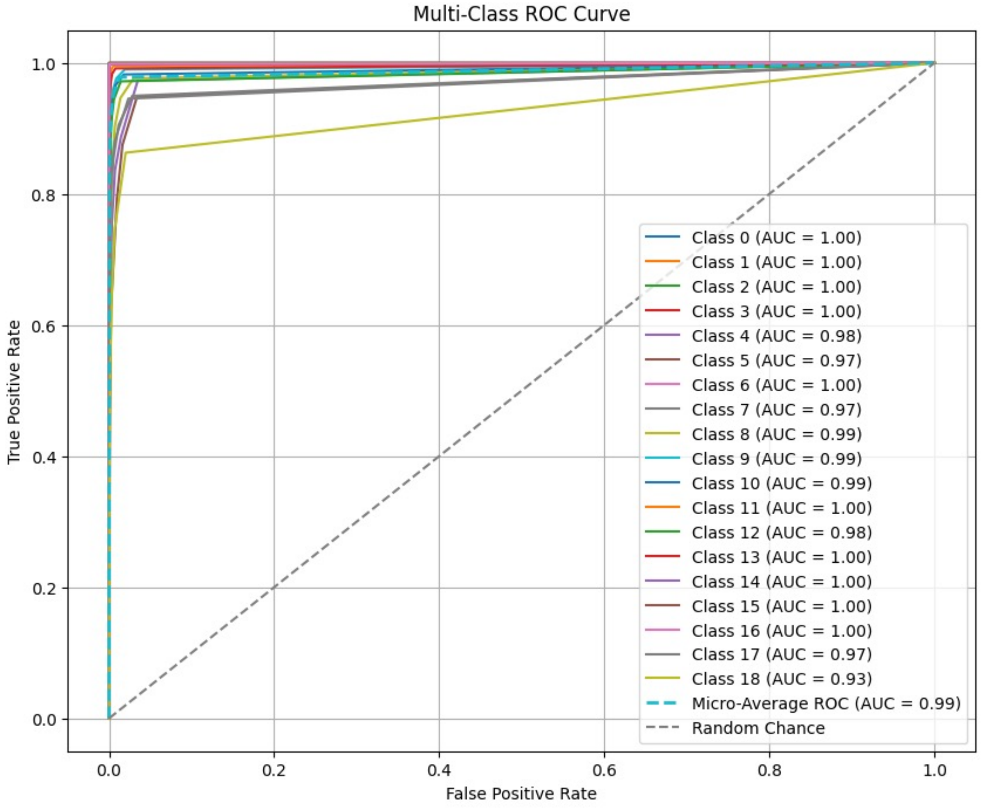
	precision	recall	f1-score	support
1	1.00	0.00	0.00	17850
2	0.00	0.00	0.00	17911
3	1.00	0.99	1.00	18019
4	1.00	1.00	1.00	17915
5	0.00	0.00	0.00	18010
6	0.00	0.00	0.00	17948
7	0.00	0.00	0.00	17989
8	0.75	0.00	0.00	18197
9	0.91	0.00	0.00	18064
10	0.29	0.83	0.43	17990
11	0.00	0.00	0.00	18028
12	0.73	0.08	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

MODEL PERFORMANCE

KNN

Classification Report:

	precision	recall	f1-score	support
1	1.00	1.00	1.00	11913
2	1.00	1.00	1.00	11946
3	1.00	1.00	1.00	12037
4	1.00	1.00	1.00	12142
5	0.85	0.87	0.86	12006
6	0.83	0.92	0.88	11999
7	0.98	0.99	0.98	12299
8	0.85	0.83	0.84	11926
9	0.87	0.90	0.88	11960
10	0.85	0.89	0.87	11999
11	0.87	0.86	0.86	12078
12	0.89	0.88	0.89	11973
13	0.86	0.90	0.88	12081
14	0.88	0.90	0.89	11950
15	0.98	0.99	0.98	11941
16	0.95	0.96	0.96	11962
17	0.99	1.00	0.99	11889
18	0.84	0.76	0.80	12097
19	0.73	0.59	0.65	11802
accuracy			0.91	228000
macro avg	0.91	0.91	0.91	228000
weighted avg	0.91	0.91	0.91	228000



Without Transformation:	67%
Dropping Highly correlated features:	84%
Implementation PCA:	91%

CONCLUSION

- KNN demonstrated superior performance with 91% 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