

Home

About

Contact



Predict, Play, Win:

Machine Learning in Competitive Volleyball

Start Slide



Applied Physics 157

Bendejo, Kevin & Pestaño, Joshua Omar

01

02

03

Background

Physical skill alone is no longer enough to determine success in professional sports. Data-driven insights lead to more informed in-game decisions. Preparation improves tactical execution, player confidence, and team coordination.

Sports analysis is a key strategic element in success.

Background

In 2024, Xia et. al. from the Department of Computer Science, University of California published a study titled

VREN: Volleyball Rally Dataset with Expression
Notation Language

VREN – Volleyball Rally Expression Notation

Background

VREN: Volleyball Rally Dataset with Expression Notation Language

This paper has two main objectives:

1. Propose a descriptive volleyball language to describe a volleyball rally process (VREN) in a formatted way.
2. Create a rich volleyball dataset (VREN dataset) using the VREN language.

The VREN dataset

The VREN dataset was obtained from real-world men's volleyball tournaments at the national team and NCAA levels.

Data was obtained from game videos using manual annotation of volleyball coaches and experts using the VREN language.

The dataset contains 1,632 rallies and 12,112 action features selected from 2019-2021 NCAA Division 1 Big West Conference.

Excerpt from the VREN dataset

	A	B	C	D	E	F	G	H
1	rally	round	team	receive_location	digger_location	pass_land_location	hitter_location	hit_land_location
2	1	1	b		4	4	13	13
3	1	2	a		7	6	3	11
4	1	3	b		6	6	21	25
5	1	4	a		13	13	13	15
6	1	5	b		9	9	3	15
7	1	6	a		4	4	3	12
8	1	7	b		10	10	3	15
9	1	8	a		8	6	8	11
10	2	1	b		3	3	13	14
11	3	1	a		2	2	12	11
12	3	2	b		12	12	12	13
13	4	1	b					
14	5	1	a		7	8	13	14
15	6	1	b		4	3	26	15
16	7	1	b		5	5	7	15
17	8	1	a		3	3	12	15
18	9	1	b		8	8	8	15
19	10	1	b		9	8	13	14
20	11	1	a		2	2	12	14

Motivations

With the VREN dataset, we are interested to know which features (i.e. “receive_location” and “digger_location”) best predict a successful attack.

Moreover, since the paper by Xia et. al. used the dataset to predict the “hitting type” and “set location”, we are also interested in developing a machine learning model that will predict these two categories.

Our findings will be useful in volleyball training, sports analysis, and making pre-match team preparations as well as split second in-game decisions.

Objectives

1. To apply basic machine learning techniques to determine the feature within the VREN dataset that most accurately predicts a successful volleyball attack, as well as the hitting type and set location.
2. To use each column in the VREN dataset to train a model that will predict a specific target column, and to compare the resulting accuracies obtained from each column.
3. To select the ten features with the highest predictive accuracy from Objective 2 and evaluate all possible feature combinations to determine which combination yields the machine learning model with the highest overall accuracy.

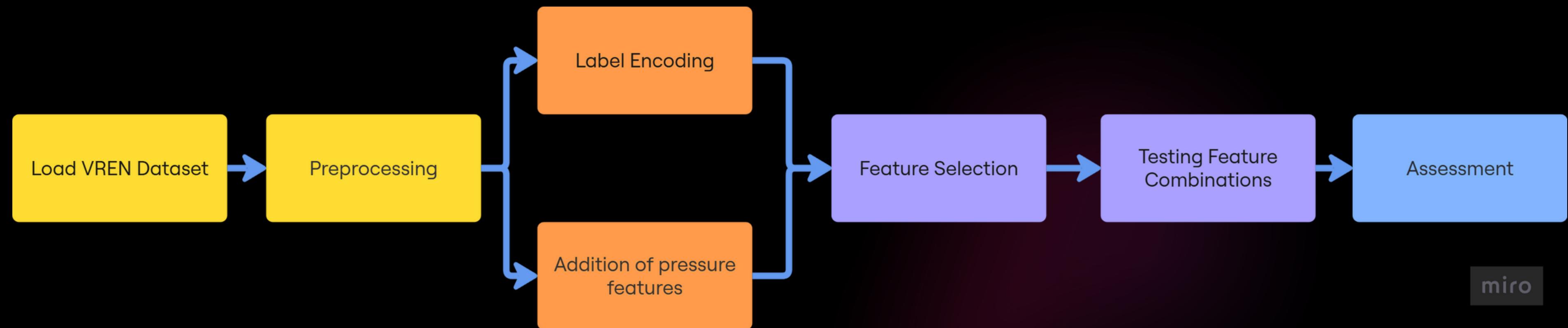
The Algorithm

An ensemble-based classifier was used to create a model that will predict successful attacks. Using an ensemble-based classifier prevents overfitting.

We used *histgradientboostingclassifier* from `sklearn.ensemble` because of the following reasons:

1. It is optimized for speed and memory so it is good for large datasets such as the VREN dataset.
2. Uses histogram-based binning to speed up training time.
3. Has native support for missing values.

The Algorithm



Results

Target	Features	Accuracy	Reference Accuracy
Successful Attack	"receive_location", "hit_land_location", "set_type", "hit_type", "block_touch", "pass_quality"	72.88%	----
Hit Type	"hitter_location", "hit_land_location", "pass_rating", "end_of_rally", "num_blockers", "block_touch"	72.33%	College Level: 71.28% Pro Level: 73.63%
Set Location	'block_touch', 'hit_type', 'num_blockers', 'pass_rating'	50.68%	College Level: 54.65% Pro Level: 51.65%

Takeaways

- Simple machine learning techniques can be used to predict volleyball variables with reasonable accuracy.
- Multiple features improve accuracy, but only up to a certain number of features.
- Since a combination of features yielded higher prediction accuracy than using any single feature, the results imply that volleyball attack success, hitting type, and set location are influenced by multiple interrelated factors.
- Results support that the VREN dataset contains valuable patterns which, together with machine learning techniques, can be used to make informed decisions in training and matches.

Appendix A: Code Snippet

```
# Load the dataset
df = pd.read_csv("processed_data.csv")

X = df[["block_touch"]].astype("category")
y = df["end_of_rally"]
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

# Create and train the model using default parameters
model = HistGradientBoostingClassifier()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred)

# Print results
print(f"Accuracy: {accuracy:.4f}")
print("Classification Report:")
print(report)
```

Appendix B: Predicting capability of each feature

Column	Feature	Accuracy (%)
M	hit_type	66.58
H	hit_land_location	61.92
N	num_blockers	61.37
AE	pass_quality	58.63
J	set_type	58.08
I	pass_rating	57.62
O	block_touch	57.53
X	Set_Number	57.53
AD	streak_B	57.53
D	receive_location	57.26
W	Score_TeamB	56.99
AA	points_to_win_B	56.99
G	hitter_location	56.44
V	Score_TeamA	56.44
Z	points_to_win_A	56.44
AC	streak_A	56.16
AG	set_loc	53.97

Column	Feature	Accuracy (%)
D	receive_location	65.75
G	hitter_location	64.11
H	hit_land_location	65.48
I	pass_rating	64.66
J	set_type	65.48
U	end_of_rally	66.58
N	num_blockers	67.4
O	block_touch	64.66
V	Score_TeamA	63.01
W	Score_TeamB	64.11
X	Set_Number	64.83
Z	points_to_win_A	66.03
AA	points_to_win_B	68.77
AC	streak_A	63.56
AD	streak_B	66.03
AE	pass_quality	64.93
AG	set_loc	64.38

Column	Feature	Accuracy (%)
G	hitter_location	90.41
N	num_blockers	47.12
O	block_touch	46.03
D	receive_location	45.21
M	hit_type	45.21
AD	streak_B	45.21
H	hit_land_location	44.93
I	pass_rating	44.93
Z	points_to_win_A	44.66
AA	points_to_win_B	43.84
X	Set_Number	43.56
AC	streak_A	42.47
W	Score_TeamB	42.19
V	Score_TeamA	41.92
AE	pass_quality	41.64
U	end_of_rally	39.73
J	set_type	37.81