

Configuration Manual

MSc Research Project
MSc Data Analytics

Megh Vijay Shirke
Student ID:x23302682

School of Computing
National College of Ireland

Supervisor: Dr. Muhammad Asif Razzaq

**National College of Ireland
Project Submission Sheet
School of Computing**



Student Name:	Megh Vijay Shirke
Student ID:	x23302682
Programme:	MSc Data Analytics
Year:	2025-26
Module:	MSc Research Project
Supervisor:	Dr. Muhammad Asif Razzaq
Submission Due Date:	11/12/2025
Project Title:	Configuration Manual
Word Count:	779
Page Count:	4

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Megh Vijay Shirke
Date:	11th December 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission , to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project , both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Megh Vijay Shirke
x23302682

This report details the software tools, libraries, and configuration settings needed to accurately reproduce the experimental environment used in the Cricket Shot Prediction Using Match Context Data with an XGBoost Classifier project.

1 Operating System

- Windows 10/11 (64-bit)
- Ubuntu 20.04+ or any Linux Distribution with Python support
- macOS

2 Hardware Requirements

- CPU: Intel i5/i7 or AMD equivalent
- RAM: Minimum 8 GB (16 GB recommended)
- Storage: At least 5 GB free space for datasets and model checkpoints

3 Programming Language

Python Version: 3.10 or higher (tested with Python 3.12.7)

4 Dataset Configuration

The dataset file must be named Shot_Extraction.csv and placed in the working directory.

Important Columns:

- shot_type
- ball_length
- ball_line
- outcome
- batsman

- bowler
- Commentary

5 Required Python Libraries

Install all dependencies via pip

Table 1: Placeholder Caption

Library	Version	Usage in Project
pandas	2.2.2	Data manipulation, cleaning, and handling datasets
numpy	1.26.4	Numerical computations, array operations,
matplotlib	3.9.2	Visualization of results, box plot, violin plot and bar plots
seaborn	0.13.2	Enhanced statistical data visualization
scikit-learn (sklearn)	1.3.0	Model training, feature extraction (TF-IDF), train-test splitting, label encoding, Classification metrics, One Hot Encoding
xgboost	2.1.1	XGBoost is a fast, scalable, and highly accurate machine-learning library that uses gradient-boosted decision trees to deliver top performance on structured/tabular data.
imblearn	0.12.3	Python library for managing and resampling imbalanced datasets in machine learning.
scipy.sparse	1.13.1	provides tools for handling and computing with sparse matrices

6 Model Configurations

- **SMOTE**
random_state=42
k_neighbors=1
- **Vectorisation:** TfidfVectorizer(max_features=1000
ngram_range=(1,2)
stop_words='english')
- **XGBoost Classifier:**
n_estimators=300,
learning_rate=0.1
max_depth=6
subsample=0.8
colsample_bytree=0.8
random_state=42
eval_metric='mlogloss'

7 Context-Aware Shot Prediction Performance Evaluation Example

Commentary: Short ball on middle stump, batsman swivels and pulls it behind square.

ball_length: short

ball_line: middle stump

bowler: Shaheen Shah Afridi

batsman: Abhishek Sharma

batting_team: India

venue: MCG

city: Melbourne

match_phase: Middle

over: 10

ball_in_over: 3

runs_batter: 1

runs_extras: 0

runs_total: 75

fatigue_index: 10/20

match_situation_index: 0.42

8 Results

Expected results are in the given form:

===== Prediction Output =====

Predicted Shot : pull

Confidence Score : 0.9966

Top-3 Probabilities : ‘pull’: 0.9966, ‘hook’: 0.0013, ‘miss’: 0.0011

References

- Achanta, A. and Boina, R. (2023). Advanced techniques in python for effective data visualization, *International Journal of Science and Research (IJSR)* **12**(12): 1919–1926.
- Banerjee, D. et al. (2025). Advancements in cricket analytics: Shot classification using deep learning techniques, *2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*, IEEE, pp. 1153–1158.
- Cricbuzz (n.d.). <https://www.cricbuzz.com>.
- Cricsheet (n.d.). <https://cricsheet.org/>.
- Das, D., Saikia, H., Bhattacharjee, D. and Kushvaha, B. (2022). On estimating shot selection by a batsman in twenty20 cricket: A probabilistic approach, *Communications in Statistics: Case Studies, Data Analysis and Applications* **8**(2): 354–367.
- Gürpınar-Morgan, W., Dinsdale, D., Gallagher, J., Cherukumudi, A. and Lucey, P. (2021). You cannot do that ben stokes: dynamically predicting shot type in cricket using a personalized deep neural network, *arXiv preprint arXiv:2102.01952*.

- Jayalath, K. P. (2018). A machine learning approach to analyze odi cricket predictors, *Journal of Sports Analytics* **4**(1): 73–84.
- Mehta, S., Kumar, A., Dogra, A. and Hariharan, S. (2024). The art of the stroke: Machine learning insights into cricket shot execution with convolutional neural networks and svm, *2024 2nd World Conference on Communication & Computing (WCONF)*, IEEE, pp. 1–6.
- Ramraj, S., Uzir, N., Sunil, R. and Banerjee, S. (2016). Experimenting xgboost algorithm for prediction and classification of different datasets, *International Journal of Control Theory and Applications* **9**(40): 651–662.
- Rehman, Z. U., Iqbal, M. M., Safwan, H. and Iqbal, J. (2022). Predict the match outcome in cricket matches using machine learning, *Manchester Journal of Artificial Intelligence and Applied Sciences* **3**(1).
- Sahoo, D. K., Deyar, D. U. and Gupta, D. (2024). Prediction of shot selection of a batsman in the death over of a t20 cricket match using machine learning, *2024 IEEE 9th international conference for convergence in technology (I2CT)*, IEEE, pp. 1–7.
- Sen, A., Deb, K., Dhar, P. K. and Koshiba, T. (2021). Cricshotclassify: an approach to classifying batting shots from cricket videos using a convolutional neural network and gated recurrent unit, *Sensors* **21**(8): 2846.
- Tyagi, A., Kaur, A., Kamboj, A., Chaulia, C., Mohan, G. and Singh, M. (2024). Xgboosting cricket: Enhancing predictive modeling for twenty20 match results using machine learning and statistical techniques, *SN Computer Science* **5**(8): 1036.
- Wei, X., Lucey, P., Morgan, S. and Sridharan, S. (2016). Forecasting the next shot location in tennis using fine-grained spatiotemporal tracking data, *IEEE Transactions on Knowledge and Data Engineering* **28**(11): 2988–2997.