PRESIDENCY UNIVERSITY
Private University Estd. in Karnataka State by Act No. 41 of 2013
Itgalpura, Rajankunte, Yelahanka, Bengaluru – 560064

PII Sentinel: A Specialized Tool for Scrutinizing Documents for Official Identifiers

A PROJECT REPORT

Submitted by

ADITYA SAHANI- 20221CBC0023

GIRIDHAR- 20221CBC0018

AFNAN PASHA- 20221CBC0012

Under the guidance of,

Ms. SUMA N G

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING, (BLOCKCHAIN)

PRESIDENCY UNIVERSITY

BENGALURU

DECEMBER 2025

PRESIDENCY UNIVERSITY

Private University Estd. in Karnataka State by Act No. 41 of 2013
Itgalpura, Rajankunte, Yelahanka, Bengaluru – 560064

PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

Certified that this report " ***PII Sentinel*** " is a bonafide work of "**ADITYA SAHANI (20221CBC0023), GIRIDHAR (20221CBC0018), AFNAN PASHA (20221CBC0012)**", who have successfully carried out the project work and submitted the report for partial fulfilment of the requirements for the award of the degree of **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE AND ENGINEERING (BLOCKCHAIN)** during **2025-26**.

28/11/2025

Ms. Suma N G
Project Guide
PSCS
Presidency University

2/12/25

Dr. H M Manjula
Program Project Coordinator
PSCS
Presidency University

2/12/25

Dr. Sampath A K
Dr. Geetha A
School Project Coordinators
PSCS
Presidency University

Dr. S Pravithraja
Head of the Department
PSCS
Presidency University

Dr. Shakkeera L
Associate Dean
PSCS
Presidency University

Dr. Duraipandian N
Dean
PSCS & PSIS
Presidency University

Examiners

Sl. no.	Name	Signature	Date
1	Shana Praveen		3/12/25
2	Bhiram Sankar.		02/12/2025

PRESIDENCY UNIVERSITY

PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

We the students of final year B. Tech **in COMPUTER SCIENCE AND ENGINEERING, BLOCKCHAIN** at Presidency University, Bengaluru, named **ADITYA SAHANI, GIRIDHAR, AFNAN PASHA**, hereby declare that the project work titled **" *PII Sentinel* "** has been independently carried out by us and submitted in partial fulfilment for the award of the degree of B.Tech in **COMPUTER SCIENCE AND ENGINEERING(BLOCKCHAIN)** during the academic year of 2025-26. Further, the matter embodied in the project has not been submitted previously by anybody for the award of any Degree to any other institution.

ADITYA SAHANI	- 20221CBC0023	SIGNATURE: Aditya
GIRIDHAR	- 20221CBC0018	SIGNATURE: Giridhar
AFNAN PASHA	- 20221CBC0012	SIGNATURE: Afnan Pasha

PLACE: BENGALURU

DATE: 26-11-2025

ACKNOWLEDGEMENT

For completing this project work, we have received the support and the guidance from many people whom I would like to mention with deep sense of gratitude and indebtedness. We extend our gratitude to our beloved **Chancellor, Vice Chancellor, Pro-Vice Chancellor and Registrar** for their support and encouragement in completion of the project.

I would like to sincerely thank my internal guide **Ms Suma N G, Assistant Professor**, Presidency School of Computer Science and Engineering, Presidency University, for his moral support, motivation, timely guidance and encouragement provided to us during the period of our project work.

I am also thankful to **Dr. S Pravithraja, Professor, Head of the Department, Presidency School of Computer Science and Engineering** Presidency University, for his mentorship and encouragement.

We express our cordial thanks to **Dr. Duraipandian N**, Dean PSCS & PSIS, **Dr. Shakkeera L**, Associate Dean, Presidency School of computer Science and Engineering and the Management of Presidency University for providing the required facilities and intellectually stimulating environment that aided in the completion of my project work.

We are grateful to **Dr. Sampath A K, and Dr. Geetha A, PSCS** Project School Coordinators, **Dr. H M Manjula, Program Project Coordinator**, Presidency School of Computer Science and Engineering, or facilitating problem statements, coordinating reviews, monitoring progress, and providing their valuable support and guidance.

We are also grateful to Teaching and Non-Teaching staff of Presidency School of Computer Science and Engineering and also staff from other departments who have extended their valuable help and cooperation.

AFNAN PASHA
ADITYA SAHANI
GIRIDHAR

ABSTRACT

The proliferation of digital documents containing Personally Identifiable Information (PII) has created significant privacy concerns, particularly with the implementation of India's Digital Personal Data Protection Act, 2023. This project addresses the critical need for specialized tools to detect official Indian identifiers in digital documents by developing PII Sentinel, a web-based application that combines advanced pattern recognition with contextual analysis.

The system employs a hybrid detection methodology integrating Regular Expressions for structured pattern matching and Named Entity Recognition using spaCy for contextual validation. The architecture features a React.js frontend, Flask backend, and PostgreSQL database, with Tesseract OCR processing scanned documents. The implementation follows an Agile methodology with five development sprints covering requirements analysis, system design, implementation, testing, and deployment.

Experimental evaluation on 537 documents containing 2,156 PII instances demonstrated 94.2% recall and 93.9% precision, significantly outperforming baseline approaches. The system processes text-based documents in under 3 seconds and scanned documents in approximately 5 seconds, supporting batch operations and real-time monitoring. User testing revealed high satisfaction scores (4.6/5.0 for ease of use) and identified valuable enhancements for future iterations.

PII Sentinel represents a substantial contribution to data privacy protection by providing organizations with an affordable, specialized solution for Indian PII detection. The system successfully bridges the gap between expensive enterprise solutions and limited opensource alternatives, enabling compliance with data protection regulations while maintaining operational efficiency.

Keywords: PII Detection, Data Privacy, Indian Identifiers, Named Entity Recognition, Hybrid Approach, Document Analysis, DPDP Act Compliance.