**A Minor Project Report on**

**Bias Detection in Hate Speech**

Report submitted to National Institute of Technology, Patna by

Archana Kumari - 1906011

Priyanshu Raj - 1906012

Garge Archana Atul - 1906124

Project done under the Supervision of  
**Dr. Jyoti Prakash Singh**  
Assistant Professor CSE Department, NIT Patna



**Department of Computer Science and Engineering**

**National Institute of Technology, Patna**



**CERTIFICATE**

This is to certify that **Archana Kumari (1906011), Priyanshu Raj (1906012), Garge Archana Atul (1906124)** have carried out the project entitled "**Gender Bias Detection in Hate Speech Classifier**" as their 6th semester Minor Project-I (CS6490) under my supervision.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Dr. Jyoti Prakash Singh**  
Assistant Professor  
 Computer Science & Engineering  
 National Institute of Technology Patna



**Declaration**

We hereby declare that this project work for Minor Project-I (CS6490) entitled "Gender Bias Detection in Hate Speech Classifier" has been carried out by us under the supervision of Dr Jyoti Prakash Singh, Assistant Professor, Department of Computer Science and Engineering, NIT Patna. No part of this project has been submitted for the award degree or diploma to any other Institute.

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No.** | **Name** | **Roll No.** | **Signature** |
| 1 | Archana Kumari | 1906011 |  |
| 2 | Priyanshu Raj | 1906012 |  |
| 3 | Garge Archana Atul | 1906124 |  |

Place: Patna   
Date: 19th May 2022

**ACKNOWLEDGEMENT**

We hereby take the privilege to express our gratitude to all the people who were directly or indirectly involved in the execution of this work, without whom this project would not have been a success. We extend our deep gratitude, respect and obligation to our project supervisor, Dr. Jyoti Prakash Singh, Assistant Professor, Department of Computer Science and Engineering, for his timely suggestions and encouragement. Our heartiest thank to our classmates who have supported us in all possible ways. Words are inadequate to express our gratitude to our parents and friends who have been supportive all the time. We would also like to thank our institution and the faculty members without whom this project would have been a distant reality.

**Archana Kumari (1906011)  
 Priyanshu Raj (1906012)  
 Garge Archana Atul (1906124)**

**Abstract**

Gender Bias in hate speech relates to expressions which spread, incite, promote or justify hatred based on sex. Some groups of women are particularly targeted by sexist hate speech (notably young women, women in the media or women politicians), but every woman and girl is a potential target for online and offline sexist hate speech.  The increasing availability and use of Internet and social platforms have contributed to growing occurrences of sexist hate speech.In our method, we have tried to detect mitigate bias in hate speech. Our method provides 74% accuracy with deep learning models and 78% accuracy with gradient boosting.