FYP Proposal
Privacy Policies and Cyber Laws

Supervisor:
Dr. Omer Beg

Signature:

_____

Group Members:
Tehreem Javed i16-0086
Ayesha Qamar i16-0104

FYP Short Title:

Cyber Compliance

Complete Title:

Privacy Policies and Cyber Law

Type of Project:

Research and Development

Group information:

Name	Roll number	CGPA	e-mail
Tehreem Javed	16i-0086	3.70	i160086@nu.edu.pk
Ayesha Qamar	16i-0104	3.88	i160104@nu.edu.pk

Supervisor:

Dr. Mirza Omer Beg

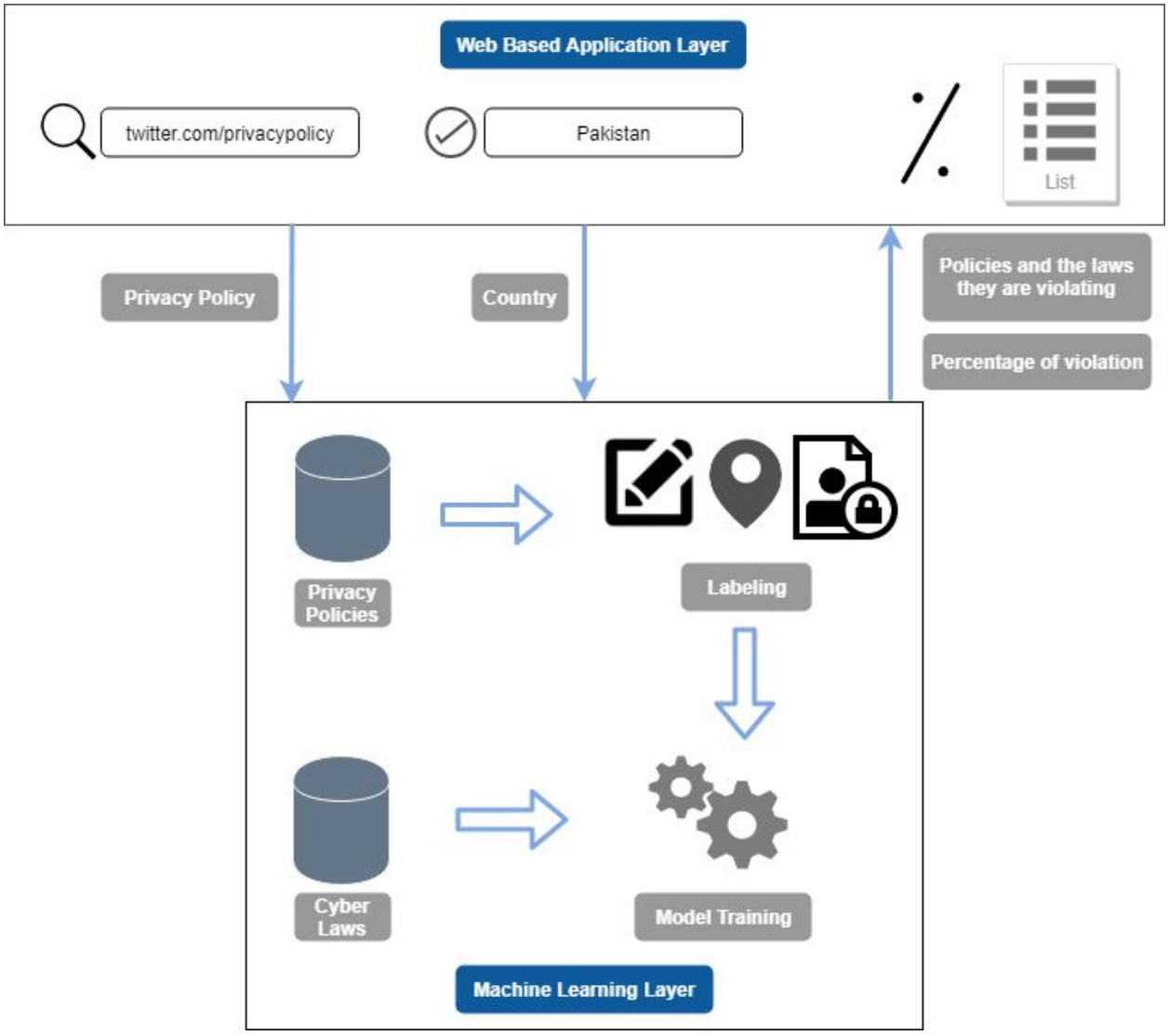
Project Overview:

Privacy policies and cyber laws regulating these policies are both highly extensive and full of legal jargon. In fact, it is estimated that about 201 hours on average are needed by any average user just to read all the privacy policies encountered in a year [1]. As a result, consumers don't fully understand what they are signing up for [2] and often do not know whether the policies that they are agreeing to are infringing on their legal rights.

Moreover, a company's legal department spends hours to review its privacy policies to see if it is compatible with a given country's laws. This is a rigorous process because each country has its own cyber laws and also because with the upsurge of Internet of things there has been an escalation in the number and complexity of privacy policies themselves [3].

In recent times in the field of nlp, work has been done on privacy policies but none that caters to the problem of verifying if a given privacy policy adheres to the laws of a given country or state. A possible solution is to create a system powered by machine learning to review a privacy policy

and see if it is in accordance to the cyber laws of the country (or countries) and identify any areas where a violation between them is detected.



Motivation:

A reliable privacy policy validation tool can be of value to companies, consumers and regulators alike. Companies can use it to help in privacy policy modeling during product launches. It can also be of help when introducing the product to a foreign country as it will significantly ease the process of verifying the policies in light of the new cyber law. Consumers can avail the service to better educate themselves regarding the policies they are agreeing to as well as their legal rights without having to go through an ambiguous legal document. Departments or organizations dedicated to cyber security can also use this service to find loopholes in any given policy and take relevant action.

Presently, work has been done to analyse privacy policies and present them to the users in a compact and easy-to-understand way. *Prolisis* is a privacy policy analysis tool which has the ability to find and present what information is being collected and why, using deep learning. It has an accuracy of 88.4% [4]. In addition, *Automated analysis of privacy requirements for mobile apps* checks the policies of an app against its practices to find conflict using automated analysis [5]. However, none of these works map policies to cyber laws to find if the policies comply to them.

Objectives:

- Gathering privacy policies from different applications and websites for training and testing data.
- Building correlations between laws and the corresponding policies.
- Verifying if the policy is in accordance to its related law.
- Implementing semi unsupervised learning as current datasets such as OPP-115 [6] and APP-350 [7] which annotate labels to policies contain a few hundred policies only.
- Understanding the legal domain as legal knowledge is mixed with natural language in the text of both privacy policies and cyber laws

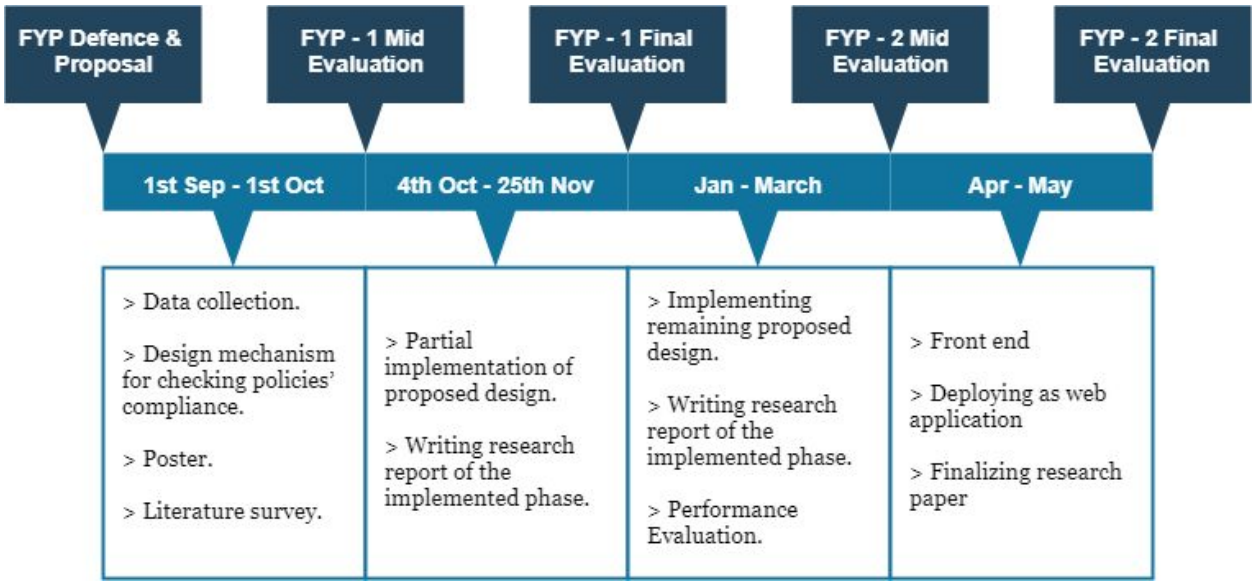
Expected Outcomes:

The end goal is to present the research through a web based interface. Users would be able to enter the link to the policy they want assessed and their location. The trained model behind is going to match the policies with the law of the country selected. It would then display back all segments of the policy which are in violation, along with the law they are violating.

Project Scope:

- Applicable on privacy policies of android apps and websites
- Semi unsupervised learning
- Match policies with various countries' laws
- Provide a web based interface for users.

Timeline:



Involved Technologies and skills:

Technology/ Skill	Skill Level of Ayesha Qamar	Skill Level of Tehreem Javed
Natural Language Processing	3	2
Machine Learning	6	6
Django	0	0
Python	8	8

References:

- [1] A. M. McDonald and L. F. Cranor, “The cost of reading privacy policies,” ISJLP, vol. 4, p. 543, 2008.

- [2] A. Rao, F. Schaub, N. Sadeh, A. Acquisti, and R. Kang, “Expecting the unexpected: Understanding mismatched privacy expectations online,” in Twelfth Symposium on Usable Privacy and Security (SOUPS 2016). Denver, CO: USENIX Association, 2016, pp. 77–96.

- [3]F. Schaub, R. Balebako, and L. F. Cranor, “Designing effective privacy notices and controls,” IEEE Internet Computing, vol. 21, no. 3, pp. 70–77, 2017.

- [4]Harkous, H., Fawaz, K., Lebret, R., Schaub, F., Shin, K. G., & Aberer, K. (2018). Polisis: Automated analysis and presentation of privacy policies using deep learning. Computation and Language.

- [5]S. Zimmeck, Z. Wang, L. Zou, R. Iyengar, B. Liu, F. Schaub, S. Wilson, N. Sadeh, S. M. Bellovin, and J. Reidenberg, “Automated analysis of privacy requirements for mobile apps,” in 24th Annual Network and Distributed System Security Symposium, NDSS 2017, 2017.

- [6] The creation and analysis of a website privacy policy corpus . Shomir Wilson, Florian Schaub and Aswarth Abhilash Dara. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, August 2016.

- [7] MAPS: Scaling Privacy Compliance Analysis to a Million Apps . Sebastian Zimmeck, Peter Story, Daniel Smullen. Privacy Enhancing Technologies Symposium 2019.