# AN APPROACH FOR PREDICTION OF GLOBAL TERRORISM BY USING MACHINE LEARNING

**A mini project report submitted in partial fulfillment of the requirements for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

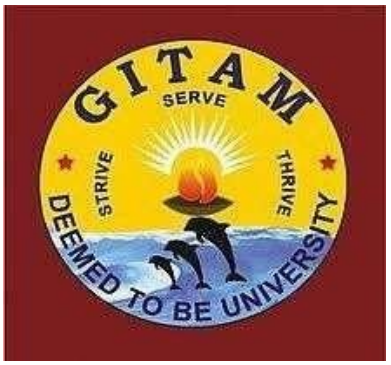
**COMPUTER SCIENCE AND ENGINEERING**

**Submitted by**


| | |
|---|---|
| **M.Laxmi Prasanna** | **221710310037** |
| **M.Rangabhagavan Reddy** | **221710310043** |
| **R Krishna Chaithanya** | **221710310057** |
| **N Rohith Reddy** | **221710310048** |


**Under the guidance of**

**Dr. Riyazuddin Y Md**





**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**GITAM**

**(Deemed to be University)**

**HYDERABAD CAMPUS**


**December - 2020**

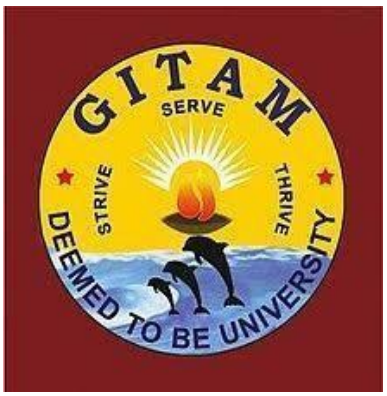
**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**GITAM SCHOOL OF TECHNOLOGY**

**GITAM**

**(Deemed to be University)**

**HYDERABAD CAMPUS**




# DECLARATION

We now declare that the mini project report entitled **"AN APPROACH FOR PREDICTION OF GLOBAL TERRORISM BY USING MACHINE LEARNING"** is an original work done in the Department of Computer Science and Engineering, GITAM School of Technology, GITAM (Deemed to be University) submitted in partial fulfillment of the requirements for the award of the degree of "Bachelor of Technology" in Computer Science and Engineering. The work had not been submitted to any other college or university to award any degree or diploma.

Date:

| Registration No(s) | Name(s) | Signature(s) |
|---|---|---|
| **221710310037** | **M.Laxmi Prasanna** | |
| **221710310043** | **M.Rangabhagavan Reddy** | |
| **221710310057** | **R Krishna Chaithanya** | |
| **221710310048** | **N.Rohith Reddy** | |

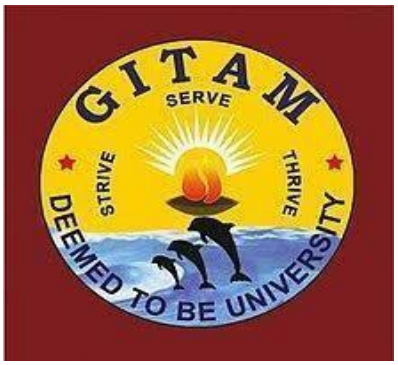
**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**GITAM SCHOOL OF TECHNOLOGY**

**GITAM**

**(Deemed to be University)**

**HYDERABAD CAMPUS**




# CERTIFICATE

This is to certify that the project report entitled "**AN APPROACH FOR PREDICTION OF GLOBAL TERRORISM BY USING MACHINE LEARNING" is** a bonafide record of work carried out by **M.Laxmi Prasanna (221710310037), M.Rangabhagavan Reddy(221710310043), R Krishna Chaithanya(221710310057), N Rohith Reddy(221710310048)** submitted in partial fulfillment of requirement for the award of degree of Bachelors of Technology in Computer Science and Engineering.

| | |
|---|---|
| **Project Guide** | **Head of the Department** |
| **Dr. Riyazuddin Y Md.** | **Dr. S. Phani Kumar** |
| **Assistant Professor** | **Professor** |

# ACKNOWLEDGEMENT

Our project would not have been successful without the help of several people. We want to thank the personalities who were part of our project in numerous ways, giving us outstanding support from the project's birth.

We are incredibly thankful to our honorable Pro-Vice-Chancellor, Prof. N.Siva Prasad, for providing the necessary infrastructure and resources for the accomplishment of our project.

We are highly indebted to Prof. N. Seetharamaiah, Principal, School of Technology, GITAM Hyderabad, for his support during the project's tenure.

We are very much obliged to our beloved Prof.S.Phani Kumar, Head of the Department of Computer Science and Engineering, School of Technology, GITAM Hyderabad, for providing the opportunity to undertake this project and encouragement in completion of this project.

We now wish to express our deep sense of gratitude to Dr. Riyazuddin Y Md, Assistant Professor, Department of Computer Science and Engineering, School of Technology, GITAM Hyderabad, for the esteemed guidance, moral support, and invaluable advice provided by him for the success of the project.

We are also thankful to all the Computer Science and Engineering department staff members who have cooperated in making our project a success. We want to thank all our parents and friends who extended their help, encouragement and moral support directly or indirectly in our project work.

| | |
|---|---|
| 221710310037 | M.Laxmi Prasanna |
| 221710310043 | M.Rangabhagavan Reddy |
| 221710310057 | R Krishna Chaithanya |
| 221710310048 | N.Rohith Reddy |

# TABLE OF CONTENTS

# ABSTRACT

It is evident that there has been enormous growth in terrorist attacks in recent years.

The idea of online terrorism has also been growing its roots in the internet world. These types of activities have been growing along with the growth in internet technology. These types of events include social media threats such as hate speeches and comments provoking terror on social media platforms such as Twitter, Facebook, etc. These activities must be prevented before it makes an impact.

In this paper, we will make various classifiers that will group and predict various terrorism activities using the k-NN algorithm and the random forest algorithm. The purpose of this project is to use the Global Terrorism Database as a dataset to detect terrorism. We will be using GTD, which stands for Global Terrorism Database, which is a publicly available database that contains information on a terrorist event far and to train a machine learning-based intelligent system to predict any future events that could bring threat to society.

# LIST OF FIGURES /TABLES /SCREENS

# 1.INTRODUCTION

## 1.1. MACHINE LEARNING

Machine Learning (ML) is the scientific study of algorithms and statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of Artificial Intelligence (AI).

## 1.2. IMPORTANCE OF MACHINE LEARNING

Consider some of the instances where machine learning is applied: the self-driving Google car, cyber fraud detection, online recommendation engines—like friend suggestions on Facebook, Netflix showcasing the movies and shows you might like, and "more items to consider" and "get yourself a little something" on Amazon—are all examples of applied machine learning. All these examples echo the vital role machine learning has begun to take in today's data-rich world.

Machines can help filter useful pieces of information that help in significant advancements, and we are already seeing how this technology is being implemented in a wide variety of industries.

With the constant evolution of the field, there has been a subsequent rise in the uses, demands, and importance of machine learning. Big data has become quite a buzzword in the last few years; that's in part due to the increased sophistication of machine learning, which helps analyze those big chunks of big data. Machine learning has also changed the way data extraction and interpretation is made by involving automatic sets of generic methods that have replaced traditional statistical technique

The process flow depicted here represents how machine learning works

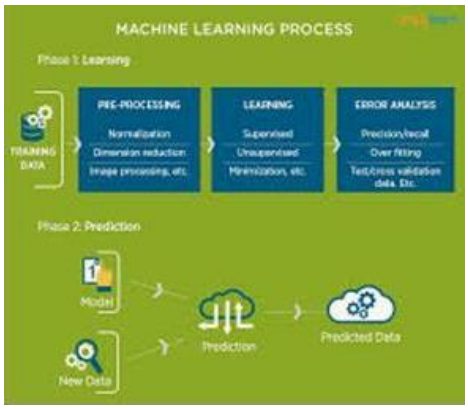



Figure 1 : The Process Flow

## 1.3. PYTHON

Python is a high-level, interpreted, interactive, and object-oriented scripting language.

Python is a general-purpose programming language that is often applied in scripting roles

Python is Interpreted: Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is like PERL and PHP.

Python is Interactive: You can sit at a Python prompt and directly interact with the interpreter to write your programs.

Python is Object-Oriented: Python supports the Object-Oriented style or technique of programming that encapsulates code within objects.

GUIDO VAN ROSSUM developed Python in the early 1990s. Its

latest version is 3.7; it is generally called as python3

## 1.4. FEATURES OF PYTHON:

i. Easy-to-learn: Python has few keywords, simple structure, and a clearly defined syntax,

ii. This allows the student to pick up the language quickly.

iii. Easy-to-read: Python code is more clearly defined and visible to the eyes.

iv. Easy-to-maintain: Python's source code is reasonably easy-to-maintaining.

v. A comprehensive standard library: Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.

vi. Portable: Python can run on a wide variety of hardware platforms and has the same interface on all platforms.

vii. Extendable: You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.

viii. Databases: Python provides interfaces to all major commercial databases.

**Terrorist attacks** are spreading at an incredible pace across the world. As per the United Nations definition of Terrorism, any action with a political goal is intended to cause death or serious bodily harm to civilians. In the last year, around 22 thousand events occurred globally, generating over 18 thousand casualties—the factors leading to terrorism change over time since they are dependent upon multiple political and social reasons. Apart from predicting the reason behind the attack, identification of the responsible agencies is also tricky. There has been a shortage of information regarding patterns of widespread terrorist behavior.

The existing analyses are either case studies or the use of quantitative methods such as regression analysis. The former is specific to certain events, while the latter approach is restricted to interviews of civilians impacted by the attack. Most of these analyses depend on weapons used for the attacks and the number of people harmed. Other types of research include the investigation of unusual patterns in individual behaviors or questioning detainees to acquire data pertaining to the attacks.

The research is focused on the correlation between terrorism and its causal factors. Existing efforts have not been good enough for prediction. Machine learning approaches can ad in predicting the likelihood of a terrorist attack, given the required data. The results of this work can help security agencies and policymakers to eradicate terrorism by taking relevant and practical measures. This paper provides an approach to analyzing terrorism region and country with the machine learning techniques and terrorism specific knowledge to fetch conclusions about terrorist behavior patterns. Through analysis of events using GTD, six supervised machine learning models (Linear Discriminant Analysis, k-Nearest Neighbours, Support Vector Machines, Decision Tree, and Logistic Regression) were built and evaluated on their performances. one of the key aims of the GTI is to examine these trends. It also aims to help inform a positive, practical debate about the future of terrorism and the required policy responses.

The GTI is based on the Global Terrorism Database (GTD), the most authoritative data source on terrorism today. The GTI produces a composite score so as to provide an ordinal ranking of countries on the impact of terrorism. The GTD is unique in that it consists of systematically and comprehensively coded data for 170,000 terrorist incidents. The GTI was developed in consultation with the Global Peace Index Expert Panel. The GTI scores each country on a scale from 0 to 10, where 0 represents no impact from terrorism, and 10 illustrates the highest measurable impact of terrorism. Countries are ranked in descending order, with the worst scores listed first in the index.

| | eventid | iyear | imonth | iday | extended | country | country_txt | region | region_txt | provstate | ... | nwound | nwoundus | nwoundte | property | ish |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 197000000001 | 1970 | 7 | 2 | 0 | 58 | Dominican Republic | 2 | Central America & Caribbean | NaN | ... | 0.0 | NaN | NaN | 0 | |
| 1 | 197000000002 | 1970 | 0 | 0 | 0 | 130 | Mexico | 1 | North America | Federal | ... | 0.0 | NaN | NaN | 0 | |
| 2 | 197001000001 | 1970 | 1 | 0 | 0 | 160 | Philippines | 5 | Southeast Asia | Tarlac | ... | 0.0 | NaN | NaN | 0 | |
| 3 | 197001000002 | 1970 | 1 | 0 | 0 | 78 | Greece | 8 | Western Europe | Attica | ... | NaN | NaN | NaN | 1 | |
| 4 | 197001000003 | 1970 | 1 | 0 | 0 | 101 | Japan | 4 | East Asia | Fukouka | ... | NaN | NaN | NaN | 1 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 181686 | 201712310022 | 2017 | 12 | 31 | 0 | 182 | Somalia | 11 | Sub-Saharan Africa | Middle Shebelle | ... | 2.0 | 0.0 | 0.0 | -9 | |
| 181687 | 201712310029 | 2017 | 12 | 31 | 0 | 200 | Syria | 10 | Middle East & North Africa | Lattakia | ... | 7.0 | 0.0 | 0.0 | 1 | |
| 181688 | 201712310030 | 2017 | 12 | 31 | 0 | 160 | Philippines | 5 | Southeast Asia | Maguindanao | ... | 0.0 | 0.0 | 0.0 | 1 | |
| 181689 | 201712310031 | 2017 | 12 | 31 | 0 | 92 | India | 6 | South Asia | Manipur | ... | 0.0 | 0.0 | 0.0 | -9 | |
| 181690 | 201712310032 | 2017 | 12 | 31 | 0 | 160 | Philippines | 5 | Southeast Asia | Maguindanao | ... | 0.0 | 0.0 | 0.0 | 0 | |

Figure 2 Global Terrorism DataBase

Defining terrorism is not a straightforward matter. There is no single, internationally accepted definition of what constitutes terrorism, and the terrorism literature abounds with competing definitions and typologies. IEP agrees with the terminology and definitions agreed to by the GTD and the National Consortium for the Study of Terrorism and Responses to Terrorism (START). The GTI, therefore, defines terrorism as 'the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation.'

This definition recognizes that terrorism is the physical act of an attack and the psychological impact on society for many years. Therefore, the index score accounts for terrorist attacks over the prior five years. In order to be included as an incident in the GTD, the act has to be 'an intentional act of violence or threat of violence by a non-state actor.' This means an incident has to meet three criteria in order for it to be counted as a terrorist act:

1. The incident must be intentional - the result of a conscious calculation on the part of a perpetrator.

2. The incident must entail some level of violence or threat of violence - including property damage as well as violence against people.

3. The perpetrators of the incidents must be sub-national actors. This database does not include acts of state terrorism. In addition to this baseline definition, two of the following three criteria have to be met in order to be included in the START database from 1997:

The violent act was aimed at attaining a political, economic, religious or social goal.

The violent act included evidence of an intention to coerce, intimidate, or convey some other message to a larger audience other than to the immediate victims.

The violent act was outside the precepts of international humanitarian law. In cases where there is insufficient information to make a definitive distinction about whether it is a terrorist incident within the confines of the definition, the database codes these incidents as 'doubt terrorism proper.' In order to only count unambiguous incidents of terrorism, this study does not include doubted incidents. It is essential to understand how incidents are counted.

According to the GTD codebook 'incidents occurring in both the same geographic and temporal point will be regarded as a single incident, but if either the time of the occurrence of the incidents or their locations are discontinuous, the events will be regarded as separate incidents. Illustrative examples from the GTD codebook are as follows:

Four truck bombs exploded nearly simultaneously in different parts of a major city. This represents four incidents.

A bomb goes off, and while police are working on the scene the next day, they are attacked by terrorists with automatic weapons. These are two separate incidents as they were not continuous given the time lag between the two events.

## 1.5. OBJECTIVES

The objective of our study is to predict the region and country of terrorist attacks. The Global Terrorism Database (GTD) is a database that is open source and includes information on terrorist events for the years 1970-2017. It includes wholesome data regarding domestic incidents, transnational, and international terrorist incidents which took place in this duration. The number of cases included is 180,000 {bombings (88,000), assassinations (19000) and kidnappings (11000)}. The parameters include the date of the incident, the month of the attack, location of the incident, and country of the incident, a region of the incident, the weapons used in the incident, nature of the target, type of attack, the number of casualties, the group or individual responsible for the incident.

## 1.6. LIMITATIONS

Global terrorism threats are difficult to defeat and pose a long-term challenge to security, especially at the individual level. But the resilience is not limited to terrorists and criminals; it is a prominent feature of contemporary states and gives them considerable protection as well. Moreover, most collective action problems limit what they can achieve. These characteristics make it extraordinarily unlikely that terrorists or other such actors will be able to undermine the national or homeland security of countries or, indeed, achieve any long-term strategic goals.

Consequently, even though terrorism poses a grave and persistent threat to personal security and a wide range of devastating attacks are possible, we argue that its threat to national security is at times poorly characterized and hence, misrepresented. The West is never likely to win a global war against terrorism or other threats, but. Like terrorists, nation-states can impose high costs on individuals or groups within the network.

Their organizational structures enable them to do so with greater efficiency and a greater degree of coordination over a more extended period of time. If diffuse terrorist groups succeed in overcoming the collective action problems, their institutionalization will generate new preferences and vulnerabilities that may make them easier to control.

## 1.7. OUTCOMES

Understand the semantic and historical development of the word "terrorism".
Analyze the concept and underpinning legal principles of international crimes of terrorism, whether at the national or international level.
Explain treaty-based crimes relevant for prosecuting acts of terrorism, whether at the national or international level.
Identify and discuss some of the reasons for, and implications of, the absence of a universally-accepted definition of terrorism at the global level.

## 1.8. PROBLEM DEFINITION

Terrorism isn’t limited to one group, geographic area, grievance, goal, method, or era. For the last century, terrorists have committed violence for all kinds of reasons: to create a new country, carry out a revolution, achieve racist goals, free animals from laboratory testing, end abortion, and more.

Analyzing terrorist groups, where they come from, why they form, and what they hope to achieve is the first step toward combating them.

We take various approaches to categorize and analyze groups to recognize patterns of behavior, locate geographic hotspots and potential targets, and understand groups' evolution over time. So we use a Machine learning algorithm to predict terrorism.

## 1.9. APPLICATIONS: -

Various Data mining techniques and machine learning algorithms like Support Vector Machine, Random Forest, Logistic Regression, etc., have been used to analyze the dataset and carry out predictions like the success of a particular attack and predict the group that carried out an attack and the effect of the external factors.

Global terrorism means the use of intentionally indiscriminate and illegal force and violence for creating terror among the masses in order to acquire some political, monetary, religious, or legal goals.

Identification of these ideologies and prediction of future attacks has proven to be of the greatest importance but is time-consuming.

This paper analyzes the historical dataset of the Global Terrorism Database and predicts the factors that might give a blow to an increase in terrorism.

A detailed comparison of each algorithm is carried out to attain the most significant results.

# 2. LITERATURE SURVEY

## 2.1 PROBLEMS WE ARE FACING DUE TO TERRORISM

**Predicting Terrorism with Machine Learning: Lessons from "Predicting Terrorism: A Machine Learning Approach"**

**AUTHORS: Basuchoudhary Atin, Bang James T**

This paper highlights how machine learning can help explain terrorism. We note that even though machine learning has a reputation for black-box prediction, in fact, it can provide deeply nuanced explanations of terrorism. Moreover, machine learning is not sensitive to the sometimes-heroic statistical assumptions necessary when parametric econometrics is applied to the study of terrorism. This increases the reliability of explanations while adding contextual nuance that captures the flavor of individualized case analysis. Nevertheless, this approach also gives us a sense of the replicability of results. We, therefore, suggest that it further expands the role of science in terrorism research.

**Terrorism's effect on Europe's center- and far-right parties**

**AUTHORS: M. Bohaneca, M.K. Borstnarb, M Robnik-Sikonja**

European far-right parties have enjoyed mixed success in the past few years. The primary elements in many of these parties' policy platforms centre on security, terrorism, and foreign persons. Naturally, these platforms are designed to attract electoral support that these actors can parlay into governing positions. Our study offers an important test to ascertain how voters respond to terrorist attacks with respect to center- and far-right parties. We contend that far-right parties are likely to benefit from terrorist attacks more than center-right parties.

The results from more than 30 European countries, spanning 1975–2013, affirm our hypothesis. The implications for partisanship, governance, and terrorism are explored in this paper as well.

**Major incidents that shaped aviation security**

**AUTHORS: Junchi Yan123, Chao Zhang2, Hongyuan Zha14, Min Gong2, Changhua Sun2, Jin Huang2, Stephen Chu2, Xiaokang Yang3**

This article is giving an overview of major incidents in civil aviation that have shaped the aviation security policies over the course of time. It begins with industry threats and security breaches (hijackings and terrorism); the countermeasures and policy decisions are giving an example of changing aviation security. The article continues with analyzing the impact of 9/11, but also the current threats to civil aviation and the international efforts in combating them.

The objective is to analyze the impact of the incidents on the evolution of aviation security and find out whether the industry has been reactive or proactive to aviation threat mitigation. This article concludes that the security methods are reactively implemented, and a proactive attitude of the stakeholders has to maintain

It course towards aviation security, as we believe the aviation will have an increasing part in the future of transportation.

**Terrorism, religion, and self-control: An unexpected connection between conservative religious commitment and terrorist efficacy**

**AUTHORS: P. Vogel, T. Klooster, V. Andrikopoulos, and M. Lungu**

Correlations between terrorism and the religious commitments of terrorist organizations and actors have been the subject of extensive scholarly investigation. 1. David Rapoport, "Fear and Trembling: Terrorism in Three Religious Traditions," American Political Science Review 78, no. 3 (1984): 658-677; Mangus Ranstorp, "Terror in the Name of Religion," Journal of International Affairs 50, no. 1 (1996): 41-62; Mark Juergensmeyer, Terrorism in the Mind of God: The Global Rise of Religious Violence (Oakland: University of California Press, 2003); Bruce Hoffman, Inside Terrorism, rev. ed. (New York: Columbia University Press, 2006), Ch.4.

View all notes Whilst the focus has often been on extreme Jihadist terrorism, other terrorist groups and individuals with religious commitments have been widely discussed, such as Baruch Goldstein's 1994 attack in Hebron, Christian Identity groups in the US, and Aum Shinrikyo in Japan. Mark Juergensmeyer, Terrorism in the Mind of God: The Global Rise of Religious Violence ibid; Bruce Hoffman, Inside Terrorism ibid. View all notes A number of theories have been advanced to explain the relationships between religious commitment and terrorism. For example, Atran has argued that many terrorists are "devoted actors" and that members of deeply conservative religions are typically devoted actors.

Scott Atran, "The Devoted Actor: Unconditional Commitment and Intractable Conflict Across Cultures," Current Anthropology 57 (2016): S192-S203.View all notes Whilst not denying that these factors may be important, and this article draws attention to a further significant impact of religion on terrorism: the surprising connection between religion and self-control.

Roy Baumeister and Julie Exline, "Virtue, Personality, and Social Relations: Self-control as the Moral Muscle", Journal of Personality 67(1999): 1165-1194; Roy Baumeister and Julie Exline, "Self-Control, Morality, and Human Strength," Journal of Social and Clinical Psychology 19 (2000): 29-42; Michael McCullough and Evan Carter, "Waiting, Tolerating, and Cooperating: Did Religion Evolve to Prop Up Humans' Self-control Abilities?" in Kathleen Vohs and Roy Baumeister (eds), Handbook of Self-regulation: Research, Theory, and Applications, 2nd ed (New York: Guilford Press, 2011); Roy Baumeister, Isabelle Bauer and Stuart Lloyd, "Choice, Freewill and Religion," Psychology of Religion and Spirituality 2, no. 2 (2010): 67-82; Sander Koole, Michael McCullough, Julius Kuhl and Peter Roelofsma, "Why Religion's Burdens are Light: From Religiosity to Implicit Self-regulation," Personality and Social Psychology Review 14 no. 1 (2010): 95-107; Michael McCullough and Evan Carter, "Waiting, Tolerating,

And Cooperating: Did Religion Evolve to Prop Up Humans' Self-control Abilities?" in Kathleen Vohs and Roy Baumeister (eds), Handbook of Self-regulation: Research, Theory, and Applications, 2nd ed (New York: Guilford Press, 2011); Kirstin Laurin, Aaron Kay and Grainne Fitzsimons, "Divergent Effects of Activating Thoughts of God on Self-regulation," Journal of Personality and Social Psychology 102 no. 1 (2012): 4-21; Connor Wood, "Ritual Well-being: Toward a Social Signaling Model of Religion and Mental Health, "Religion, Brain, and Behavior (2016) http://dx.doi.org/10.1080/2153599X.2016.1156556 View all notes Drawing on the large empirical literature establishing a link between religion (in particular deeply conservative religions) and self-control, it is hypothesized that the religious practices of religiously-inspired terrorists enhances their self-control and thus raises their efficacy, operationalized as casualties per attack.

This hypothesis will be referred to as TERS (Terrorist Efficacy, Religion, and Self-control). TERS predicts that highly conservative religious terrorist groups such as Al Qaeda and ISIS will typically have higher levels of terrorist efficacy than non-religious or moderately religious groups, and the research supports this hypothesis.

James Piazza, "Is Islamist Terrorism More Lethal? An Empirical Study of Group Ideology, Organization and Goal Structure," Terrorism and Political Violence 21 no. 1 (2009): 62-88. View all notes In addition, it is argued that TERS provides a crucial addition to Piazza's emphasis on "universal/abstract" ideologies.

Piazza, "Is Islamist Terrorism More Lethal?" (see note 5 above). View all notes Piazza remarks that left-wing terrorist groups can have universal/abstract ideologies. Case studies of atheist left-wing terrorist groups—specifically the Red Army Faction and the Red Brigades—reveal that such groups satisfy Piazza's characterization of universal/abstract groups but have casualty rates per attack very much lower than those of highly conservative religious groups such as Al Qaeda and its affiliates. It is argued that highly conservative religious convictions enhance the self-control of the latter groups, raising their efficacy relative to the former atheist groups.

## 2.2 RELATED TO TERRORISM ACTIVITIES

- **Terrorism**

Terrorism is a type of collective violence that directly impacts peace, the normal routine of a country/community, and security, and a way to generate fear in civilians using violence. The word terrorism means to frighten, which is originated from the Latin word Terre. Terrorism is a repeated action of violence carried out by an individual or group for some criminal or political reasons. Some of the causes of terrorism events are injustice like political or social injustice, religious, a belief like causing such destructive activity will be effective, illiteracy, and so on. Terrorism impacts more on the economy as it results in loss of lives, business, infrastructures, expenditure on security by the government.

- **Counter-Terrorism**

Counter-terrorism is a collection of activities that might include techniques, tactics, or strategies carried out by either government, politicians, police department, business, or military to prevent terrorism.

- **Forecasting attacks**

Predicting future attacks is a recent trend in research and not so efficient to employ because of lack of real-time data; even if the data is available, it will be noisy, i.e., it consists of unsolved attacks, the accuracy of a prediction model for such model is directly proportional to the amount of data available. In order to predict the future attack, we need to have information about past episodes like where the attack had happened, what type of weapon they have used, and the target of an attack, which contributes more to predict the kind of attack that might occur in a given location.

## 2.3. THEORETICAL FRAMEWORK

After this conceptual overview of terrorism and jihadism, the paper now moves on to what the literature has already studied and identified as key explanatory variables of the phenomenon. Since researches targeting specifically the recent phenomenon of European foreign fighters are almost inexistent and even more at the cross-national level, this chapter focuses on the studies and primary outcomes from the terrorism literature. This theoretical framework provided by several terrorist scholars and political scientists will delineate the route map towards the paper's research design. In this chapter, particular attention is dedicated to understanding which becomes a terrorist and why, at the individual and country level. The first main wave of researches on terrorism causes was carried out from the 1960s to the mid-1980s and was mainly offering theoretical arguments of psychological nature based on subjective interpretations of observations at an individual level. At that time, terrorists' behaviors and motivations were believed to be a consequence of internal psychological disorders, which demarcated them from non-terrorist. However, while some "lone wolves" may suffer from some kind of psychopathology, "thirty years of research has found psychopathology and personality disorder no more likely among terrorists than among non-terrorists from the same background." Indeed, with the

need to be more organized, effective, and important in size. Terrorist organizations often have intense screening processes in order to look for cooperative, loyal, discrete, and sociable partisans, qualities not beard by people suffering from mental illnesses.

Therefore, John Horgan points out that "despite their attractiveness, personality

traits are useless as predictors of understanding why people become terrorists". As psychological explanations were refuted by the majority of terrorist authors, political scientists and scholars started looking at external root causes, such as socio-economic variables, to explain the phenomenon at the micro-level. Indeed, from political personalities like President G. Bush to religious figures and public intellectuals, economic deprivation

Or the lack of education became easy and popular explanations of terrorism. Alberto Piazza explains that this popular and politically attractive assumption rests on a reasonable principle: “If citizens of a country are denied the means to satisfy their basic human needs, are deprived of access to reasonable economic opportunities, or are faced with glaring levels of socioeconomic inequality; they will become hopeless and enraged and may view political violence as an acceptable mean for redressing their grievances”. This correlation was also inspired by the results of studies in other political violence fields such as

Collier and Hoeffler found a positive relationship between economic variables and civil wars or Alesina and Perotti, suggesting that poor economic conditions increase the likelihood of political coups.

Nevertheless, empirical researches on terrorism have demonstrated almost unanimously that individual material deprivation and inadequate education have to be rejected to explain support or participation in terrorism. Indeed, on the contrary, those micro-level studies have even shown that terrorists tend to be drawn from well-educated, middle-class or high-income backgrounds, as illustrated by Hassan in her research on Hamas participants by Krueger and Maleckova on Hezbollah supporters, or by Sageman on al Qaeda members. In his book, Krueger explains this observation by the fact that richer and better-educated people, even if they have a higher opportunity cost of time, have better access to information and grant more importance in participating and influencing politics than lower-income and low educated people.

Therefore, elites, caring more for the cause and ideological factors than for material ones, are more often

supporters of extreme ideas and are easier to radicalize. Also, on the demand side, terrorist organizations prefer recruiting members from those socioeconomic levels in order to optimize the effectiveness of their operations and minimize costs of failure, as claimed by Benmelech and Berrebi in their study showing that better-educated terrorists are more effective in the fulfillment of complicated tasks. After this short literature review on the definition of a possible terrorist profile at the micro-level, it is even more interesting in the case of this research paper to have a look at authors that have analyzed plausible variables that could explain terrorism at the national or macro-level, i.e., why some countries suffer more from terrorism than others. In this particular field, two divergent views can be notified.

On the one hand, scholars such as Krueger & Maleckova Abadie Piazza, Krueger or Krueger & Laitin defend the view that socio-economic aspects are not useful for understanding terrorism root causes national level. Indeed, while wealthier countries are more likely to be targetted by terrorist attacks, Krueger and Laitin have shown that economic performance (GDP per capita) is statistically not a good predictor of which country extremists emerge from while controlling for political regime. In addition, the authors found no statistically significant correlation between GDP growth, the volume of international trade, education variables, or religions and terrorism in their cross-country analysis. Also, Li and Schaub did not find any correlation between FDI or portfolio investment and terrorist violence but demonstrated that economic development, on the contrary, plays a role in the reduction of terrorism in a country. Furthermore, in his

More statistically and technically advanced study, Abadie confirms the idea that GDP per capita has no significant relationship with terrorism risk, but he also demonstrated the insignificance of inequality measures (GINI) or human development indicators, which is backed by Krueger. Finally, James Piazza also rejects unemployment, poverty, or economic growth to be explanatory variables for terrorism. On the other hand, Martha Crenshaw argues that economic modernization is a significant permissive cause of terrorism by creating opportunities and vulnerabilities due to an increased complexity at all levels of society. Freytag et al. argue that socioeconomic factors play a role in the development of domestic terrorism and claim that in countries where economic growth is slow, and institutions are poor, opportunity

costs of terror are low, raising the likelihood of terrorist activities and their support from the population. They also found that human capital, i.e., education, trade openness, and GDP per capita, are positively correlated with terrorism. Those results are even more relevant when a certain level of development has been reached by the country, such as in OECD or European countries. Finally, Burgoon argues that a country's welfare efforts are negatively correlated with terrorism activities. Thus, the strengthening of social welfare policies in a

country may help to diminish domestic and international terrorism by reducing inequality, economic insecurity, poverty, and religious-political extremism. Also, he has shown that government capacity is positively correlated with terrorist attacks, but that trade openness has no relationship with terrorism incidents in a country.

In the last part of this literature review on terrorism, the paper draws the attention of the reader to some political variables that scholars have analyzed in their cross-country studies in order to explain terrorism root causes. One of the first authors that tried to explain terrorism with political variables is Martha Crenshaw. She believes that the lack of government terrorist prevention and the absence of effective security measures are key preconditions for terrorism. Also, she identifies the existence of discriminative policies or concrete grievances,

i.e., related to minority rights among a subgroup of a larger population and the lack of opportunity for political participation or political persecution and repression as key variables leading to the increase of terrorist activities in a country by creating dissatisfaction. While Krueger and Laitin as well as Krueger and Maleckova, found no explanation using economic variables, conversely, their empirical study has shown that terrorists tend to

come more from countries suffering from political oppression, i.e. where civil liberties are neglected. In the same spirit, Abadie also argues that increased political freedom has a negative impact on terrorism, after a certain point. Furthermore, in his study on captured terrorists in Iraq, Krueger found out that greater civil liberties and political rights were negatively correlated with the country of origin of the foreign fighters. The conclusion of the literature review on the determinants of terrorism is that no consensus among scholars has been met until now and that still today, it is not possible to draw an exact

terrorist profile. At the micro-level, authors have found no statistically significant relationship between economic variables or social variables and participation in terrorism. On the contrary, empirical evidence seems to favor the idea that terrorists tend to be wealthier and more educated than the average population where they come from. At the macro-level, the majority of the authors argue in the same way, i.e., that poor socioeconomic conditions are not the root for terrorism activities. Mainly, they have demonstrated that poorer countries or less-educated populations do not tend to have more people taking part in extremist or radical activities on their soil than wealthier countries and better-educated populations. Those relationships are inexistent or, even in some studies, slightly positive. As Krueger wrote, this is finally quite logical as terrorism is supposed to be a political phenomenon rather than an economic one. However, a small group of authors do not follow this view and, conversely, claim that socioeconomic conditions are important determinants for terrorism as terror is more easily generated in bad socioeconomic environments by reducing costs of terror. Finally, the only variables that seem to be somewhat explanatory for terrorism at a national level are of political nature. Indeed, studies have shown that countries where the population is more oppressed, persecuted, or repressed, thus, where civil liberties or political rights are limited, tend to generate more terrorists than real democracies. In order to finish this literature review accurately on terrorism, it is important to add that the field still has some gaps. Indeed, the literature still lacks comprehensive databases on terrorists and their background, which diminishes the quality of quantitative and qualitative researches carried out by scholars. Therefore, a particular effort should be made for accurate data collection in order to better understand the root causes of terrorism and improve the effectiveness of counter-terrorism policies considerably. In addition, while quite a lot of studies have been carried out on terrorism profile at micro-level, researches on causal explanation at cross-country level are still limited, even more on foreign jihadi fighters.

# 3.SYSTEM ANALYSIS

Machine learning algorithms have been used recently to study the different factors of terrorism because of the fact that a huge amount of labeled data is available recently. The advancements in computer technologies have been able to create many powerful computer systems to perform the required computation. In this paper, machine models are used to make predictions of different factors that lead to terrorist activities. The model is helpful for law enforcement agencies to make a prediction before an incident actually happens and potentially causes the loss of precious lives. The predicted factors are explained below.

## 3.1.FEATURE SELECTION

The National Consortium for the Study of Terrorism and Responses to Terrorism (START) has prepared a dataset known as Global Terrorism Database (GTD) (https://www.start.umd.edu/gtd). GTD contains information about terrorist activities from 1970 until 2018, including more than 181,000 different instances of terrorism. In this paper, 34 attributes (some attributes are redundant and hence discarded) are taken for the analysis.

## 3.2. PREDICTION OF DIFFERENT FACTORS OF TERRORIST ACTIVITIES:

**Suicide**: This field indicates whether the attack is suicide or not suicide. 1 = "Yes" means that the incident was a suicide attack. 0 = "No" means there is no indication that the incident was a suicide attack. The dimension of the dataset is. 90% of data is used for training (315,104 instances), and 10% is used for testing (35,012 instances). Both "Yes" and "No" classes have 175,058 instances.

**Success**: This field indicates the success of a terrorist strike. 1 = "Yes" means that the incident was successful. 0 = "No" means that the incident was not successful. The dimension of the dataset is. 90% of the dataset is taken as training (290,937 instances), and 10% is taken as a testing (32,327 instances). Each class has 161,632 instances.

**Weapon Type**: This field indicates the general type of weapon used in the incident. In the dataset, 13 different labels are used to represent a different type of weapon. These labels are explained below.

- Biological
- Chemical
- Radiological
- Left as blank

- Explosives
- Fake weapons
- Incendiary
- Melee
- Vehicle (not to include vehicle-borne explosives, i.e., car or truck bombs)
- Sabotage Equipment
- Other
- Unknown

The dimension of the dataset is. 90% of the dataset is taken as training (998,200 instances), and 10% is taken as a testing (110,912 instances). Each class has 92,426 instances.

**Region:** This field indicates 12 different regions. These regions are explained below.

- North America
- Central America and the Caribbean
- South America
- East Asia
- Southeast Asia
- South Asia
- Central Asia
- Western Europe
- Eastern Europe
- The Middle East and North Africa
- Sub-Saharan Africa
- Australasia and Oceania

The dimension of the dataset is. 90% of the dataset is taken as training (545,119 instances) and 10% as a testing (60,569 instances). Each class has 50,474 instances.

**Attack Type:** This field indicates the general method of attack and a broad class of tactics used. In the dataset, 9 different labels are given and are explained below.

- Assassination
- Armed assault
- Bombing/explosion
- Hijacking
- Hostage-taking (barricade incident)
- Hostage-taking (kidnapping)
- Facility/infrastructure attack
- Unarmed assaults
- Unknown

The dimension of the dataset is. 90% of the dataset is used for training (861,517 instances), and 10% is used for testing (95,725 instances). Each class has a 88,255 instance

## 3.3. TEXT TO NUMBERS

In the GTD dataset, some features are in text format, for instance, group name, country name, etc. It is not possible to process features with text data in machine learning. There exist multiple techniques to convert text data to numbers, e.g., TFIDF, Word2Vec, GloVe, One hot encoding, etc. In this paper, LabelEncoder class of sklearn library is used to convert non-numeric data to numeric data, as the labels are hashable and comparable to numerical labels.

## 3.4. MISSING DATA

The dataset contains many missing values, i.e., the cell does not contain any data, which results in NaN when processed by. Different interpolation techniques can be used to fill the missing data. In this paper, SimpleImputer of sklearn library is used to fill the missing data. We have replaced the missing values by mean along each column.

## 3.5. DEALING WITH UNBALANCED CLASSES

During the analysis of the dataset, it is observed that the data are not balanced in different classes. In some classes, there are more instances, while others have very few instances. Machine learning algorithms trained on unbalanced data are biased towards the classes having more instances. In order to keep the data in a balanced form, the SMOTE: Synthetic Minority Oversampling Technique presented by Chawla and later made available as a tool to be used in Python is used.

# 4.SYSTEM DESIGN

## 4.1.SYSTEM ARCHITECTURE

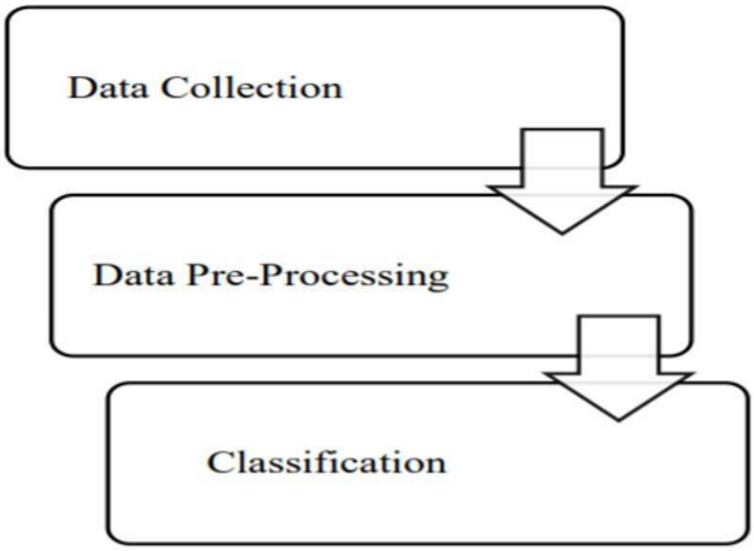



Figure 3: system architecture

### 4.1.1 DATA COLLECTION

Data collection is the process of gathering and measuring information on variables of interest in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes.

### 4.1.2. DATA PRE-PROCESSING

Data preprocessing is a data mining technique that is used to transform the raw data into a useful and efficient format. Steps Involved in Data Preprocessing:

To handle this part, data cleaning is done. It involves handling missing data, noisy data, etc.

i. **Exploratory Data Analysis:-**

Before building the model and to gain a high-level understanding of dataset features, we performed some

exploratory data and analysis.

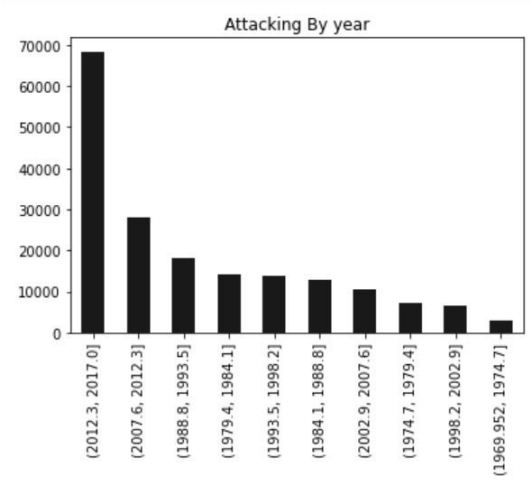



Fig. 4. Number of Yearly Terrorist Attacks

Fig. 4 depicts a significant increase in the number of terrorist attacks from 2012 and reaches a peak in 2017.

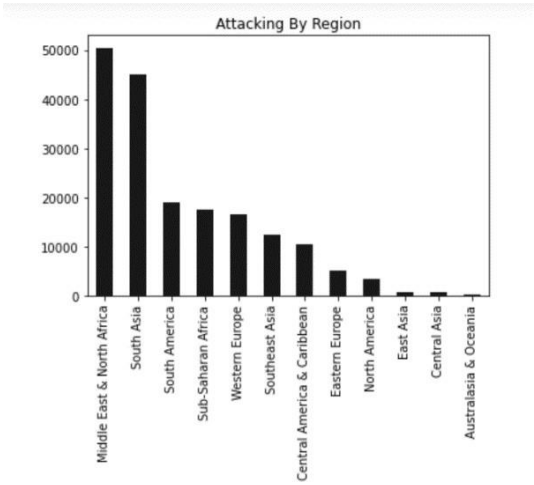



Fig. 5. Regions Attacked

The graph in Fig. 5 gives the overview of regions that are targeted by terrorists, North Africa being the top targeted region and South Asia being the second across other regions.

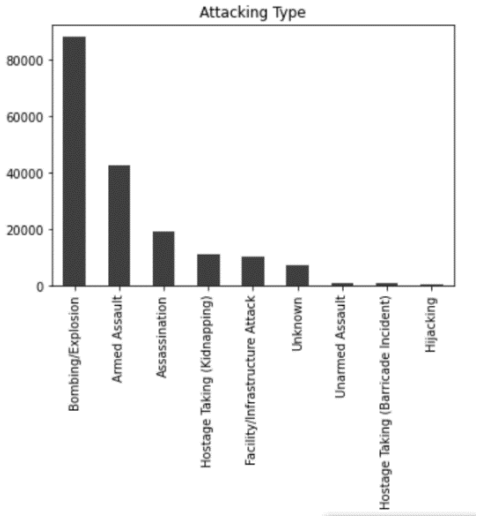



Fig. 6. Type of Attacks

The type of attacks mostly happened from 1970 to 2017 are widely conducted by Bombing /explosions (Fig. 6). It is interesting to see that armed assault and Kidnapping are the types of attacks that are widely used by terrorists after Bombing and explosions.

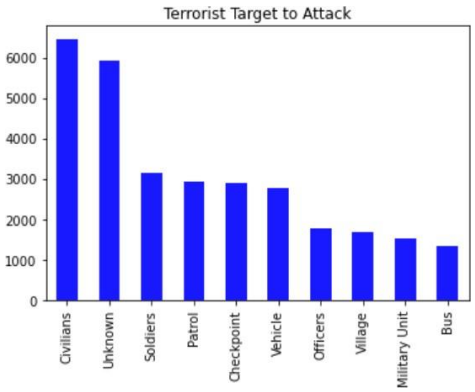



Fig. 7. Target Type

Civilians are targets among the targets, and unknown forces are the second most favorable target of the terrorists' attacks from 1970 to 2017, as depicted by Fig. 7.

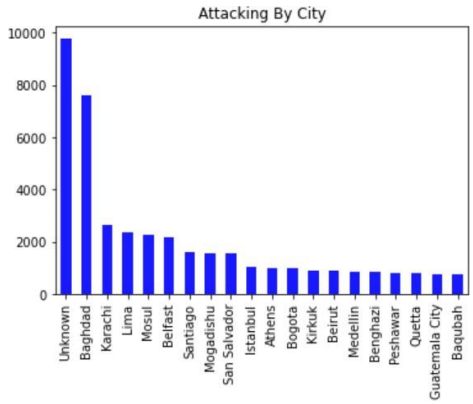



Fig. 8. Target Cities

Unknown cities have received the most terrorist attacks globally, and Baghdad comes a next number of cities that have received terrorist attacks, as may be observed from Fig. 8.

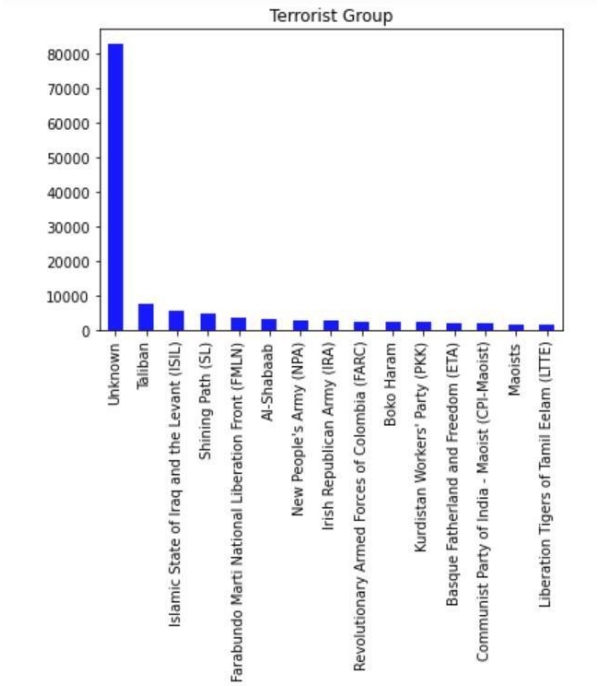



Fig. 9. Terrorist Groups

Unknown is the terrorist group which has conducted a maximum number of terrorist attacks, the Taliban being the second and ISIL in India are on the third number of top terrorist groups (Fig. 9).

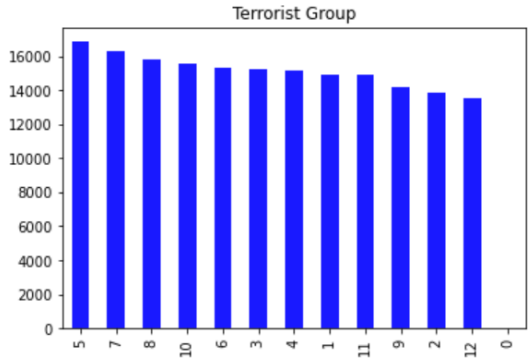



Fig. 10. Month of Attacks

May and July are the months that faced terrorist attacks most frequently from 1970 to 2017, as can be seen in Fig.10.

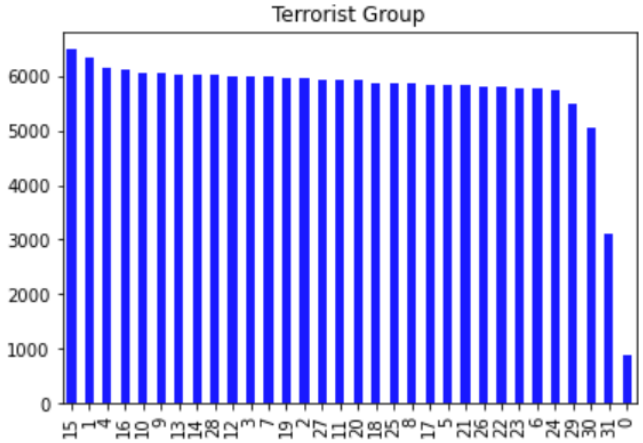



Fig. 11. Day of Month

It can be observed from Fig. 11 that the most common date of terrorist attacks is 15 th .

### 4.1.3CLASSIFICATION

In machine learning, classification refers to a predictive modeling problem where a class label is predicted for a given example of input data. Examples of classification problems include: Given an example, classify

## 4.2 REQUIREMENT ANALYSIS:

The project involved analyzing the design of few applications so as to make the application more users friendly. To do so, it was really important to keep the navigations from one screen to the other well ordered and at the same time reducing the amount of typing the user needs to do. In order to make the application more accessible, the browser version had to be chosen so that it is compatible with most of the Browsers

## 4.3INPUT AND OUT PUT DESIGN

### 4.3.1.INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation, and those steps are necessary to put transaction data into a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document, or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps, and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining privacy. Input Design considered the following things:

- What data should be given as input?
- How should the data be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when errors occur.

### 4.3.2. OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle a large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data can be performed. It also provides record viewing facilities.

3. When the data is entered, it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as to when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

### 4.3.3 OUTPUT DESIGN

Quality output is one, which meets the requirements of the end-user and presents the information clearly. In any system, results of processing are communicated to the users and to other systems through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source of information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analyzing the design computer output, they should Identify the specific output that is needed to meet the requirements.

2.Select methods for presenting the information.

3.Create a document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

Convey information about past activities, current status, or projections of the

- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

# 5.IMPLEMENTATION

## 5.1. ALGORITHMS

### 1. Logistic Regression Classification:-

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of the target or dependent variable is dichotomous, which means there would be only two possible classes.

In simple words, the dependent variable is binary in nature, having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no).

Mathematically, a logistic regression model predicts P(Y=1) as a function of X. It is one of the simplest ML algorithms that can be used for various classification problems such as spam detection, Diabetes prediction, cancer detection, etc.

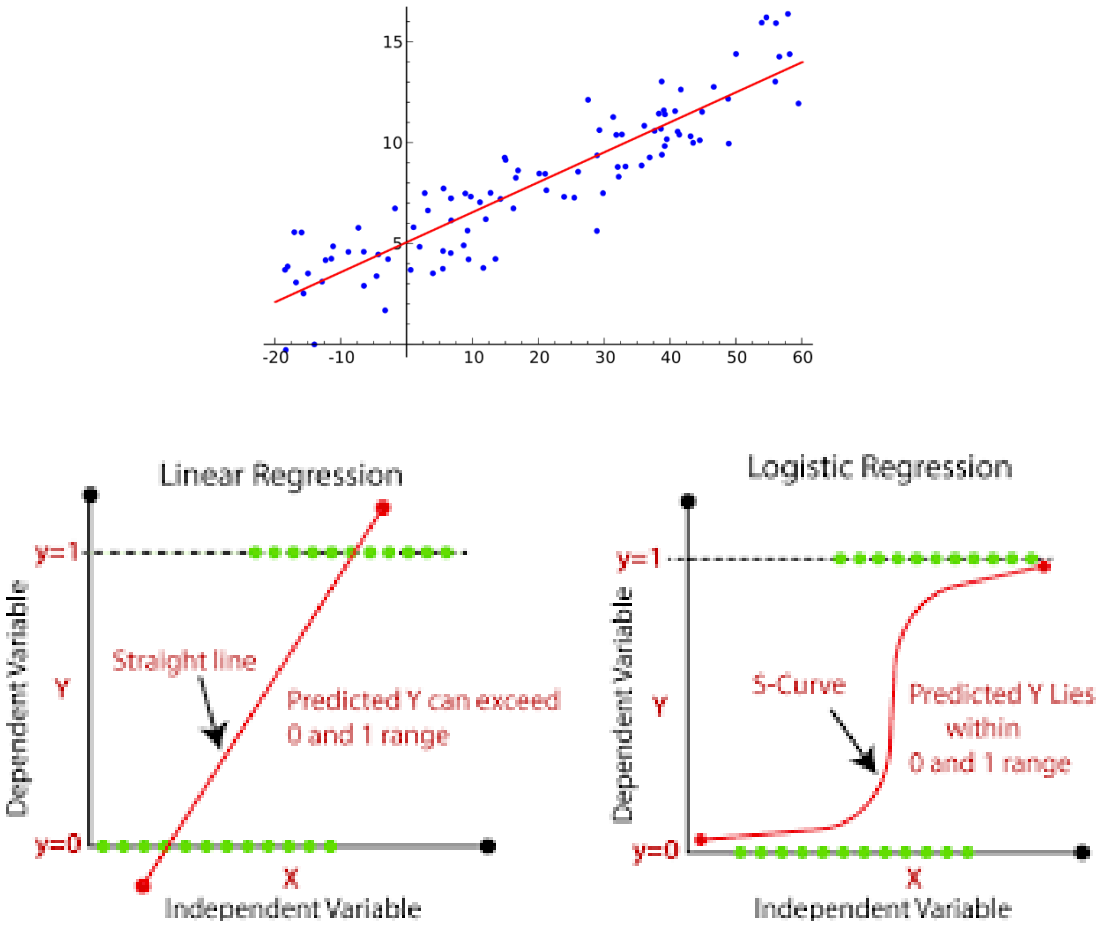



Figure 12:logistic regression

```
from sklearn.linear_model import LogisticRegression
lr = LogisticRegression()
data.dropna(axis='columns')
cols = data.columns
#print(type(cols))
train_cols = cols.drop([ 'eventid', 'iyear', 'imonth', 'iday', 'extended', 'country',
       'country_txt', 'region', 'region_txt', 'provstate', 'city', 'latitude',
       'longitude', 'specificity', 'vicinity', 'summary', 'crit1', 'crit2',
       'crit3', 'doubtterr', 'multiple',  'suicide', 'attacktype1',
       'attacktype1_txt', 'targtype1', 'targtype1_txt', 'targsubtype1',
       'targsubtype1_txt', 'corp1', 'target1', 'natlty1', 'natlty1_txt',
       'gname', 'guncertain1', 'individual', 'nperps', 'nperpcap', 'claimed',
       'weaptype1', 'weaptype1_txt', 'weapsubtype1', 'weapsubtype1_txt',
       'weapdetail',
       'nwoundte', 'property', 'ishostkid', 'dbsource', 'INT_LOG', 'INT_IDEO',
       'INT_MISC', 'INT_ANY'])
features = data[train_cols]
target = []
ans = features['success']
for val in ans=='Success':

    if val==True:
        target.append(int(True))
    else:
        target.append(int(False))

cols.drop([ 'success'])

status_replace = {
    "success" : {
        1: "Success",
        0: "Filed",
    },
}
data = data.replace(status_replace)
features = features.iloc[:, [1,2,3,4]]
features.dropna(axis='columns')
features = features.replace(0,np.NaN)
features = features.fillna(0)
lr.fit(features, target)
predictions = lr.predict(features)
```

figure 13: logistic regression algorithm code

**Types of Logistic Regression**

Generally, logistic regression means binary logistic regression having binary target variables, but there can be two more categories of target variables that can be predicted by it. Based on those number of categories, Logistic regression can be divided into the following types −

**i. Binary or Binomial**

In such a kind of classification, a dependent variable will have only two possible types, either 1 and 0. For example, these variables may represent success or failure, yes or no, win or loss, etc

### ii. Multinomial

In such a kind of classification, the dependent variable can have 3 or more possible unordered types or the types having no quantitative significance. For example, these variables may represent “Type A” or “Type B” or “Type C.”

### iii. Ordinal

In such a kind of classification, a dependent variable can have 3 or more possible ordered types or the types having a quantitative significance. For example, these variables may represent “poor” or “good,” “very good,” “Excellent,” and each category can have scores like 0,1,2,3.

## 2. Knn Algorithm

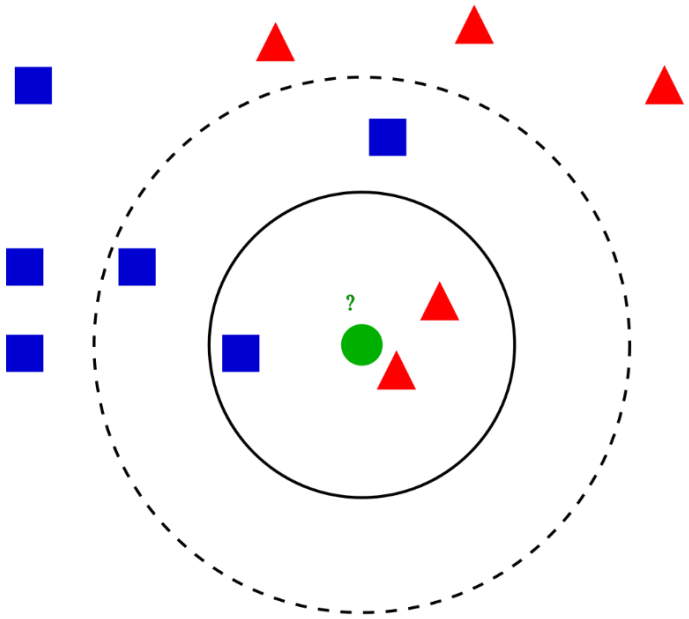



Figure 14: The test sample (green dot) should be classified either to blue squares or to red triangles. If k = 3 (solid line circle) it is assigned to the red triangles because there are 2 triangles and only 1 square inside the inner circle. If k = 5 (dashe

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on the Supervised Learning technique.

K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears; then it can be easily classified into a good suite category by using K- NN algorithm.

K-NN algorithm can be used for Regression as well as for Classification, but mostly, it is used for Classification problems.

K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data.

```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix
from sklearn import metrics
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()

data = pd.read_csv('gtd_data.csv')
data.head()
data.success.value_counts()
sns.countplot(x="success", data=data, palette="bwr")
#plt.show()

sns.countplot(x='success', data=data, palette="mako_r")
plt.xlabel("Target Success (0 = Failed, 1= Success)")
#plt.show()

'''
plt.scatter(x=data.success[data.success==1], y=data.success[(data.success==1)], c="green")
plt.scatter(x=data.success[data.success==0], y=data.success[(data.success==0)], c = 'black')
plt.legend(["Attack", "Not Attack"])
plt.xlabel("Success")
plt.ylabel("Maximum Heart Rate")
plt.show()'''



X_train, X_test, y_train, y_test =  train_test_split(features, target,test_size = 0.25, random_state= 0)
sc_X = StandardScaler()
X_train = sc_X.fit_transform(X_train)
X_test = sc_X.transform(X_test)
classifier = KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p = 2)
classifier = classifier.fit(X_train,y_train)
y_pred = classifier.predict(X_test)
#check accuracy
accuracy = metrics.accuracy_score(y_test, y_pred)
print('KNN Accuracy: ',accuracy)
```

Figure 15: knn algorithm code

It is also called a lazy learner algorithm because it does not learn from the training set immediately. Instead, it stores the dataset, and at the time of classification, it performs an action on the dataset.

KNN algorithm at the training phase just stores the dataset, and when it gets new data, then it classifies that data into a category that is much similar to the new data.

## 3. LDA (Linear Discriminant Analysis)

Linear Discriminant Analysis or LDA, is a dimensionality reduction technique. It is used as a pre-processing step in Machine Learning and applications of pattern classification. The goal of LDA is to project the features in higher dimensional space onto a lower-dimensional space in order to avoid the curse of dimensionality and also reduce resources and dimensional costs.

The original technique was developed in the year 1936 by Ronald A. Fisher and was named Linear Discriminant or Fisher's Discriminant Analysis. The original Linear Discriminant was described as a two-class technique. The multi-class version was later generalized by C.R Rao as Multiple Discriminant Analysis. They are all simply referred to as the Linear Discriminant Analysis.

```
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.tree import DecisionTreeClassifier
lda = LinearDiscriminantAnalysis()
regressor =  LinearDiscriminantAnalysis()
X_train, X_test, y_train, y_test = train_test_split(features, target, test_size = 0.3, random_state = 100)
regressor.fit(X_train, y_train)
y_pred = regressor.predict(X_test)
#print("Confusion Matrix: ",confusion_matrix(y_pred.round(),y_test))
accuracy = metrics.accuracy_score( y_pred.round(),y_test)
print("Linear Discriminant Analysis Accuracy ",accuracy)
```

Figure 16: LDA algorithm code

LDA is a supervised classification technique that is considered a part of crafting competitive machine learning models. This category of dimensionality reduction is used in areas like image recognition and predictive analysis in marketing

## 4.Decision Trees

```
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
gtd_data = pd.read_csv('gtd_data.csv')
gtd_data.dropna(axis='columns')
cols = gtd_data.columns
#print(type(cols))
train_cols = cols.drop([ 'eventid', 'iyear', 'imonth', 'iday', 'extended', 'country',
       'country_txt', 'region', 'region_txt', 'provstate', 'city', 'latitude',
       'longitude', 'specificity', 'vicinity', 'summary', 'crit1', 'crit2',
       'crit3', 'doubtterr', 'multiple',  'suicide', 'attacktype1',
       'attacktype1_txt', 'targtype1', 'targtype1_txt', 'targsubtype1',
       'targsubtype1_txt', 'corp1', 'target1', 'natlty1', 'natlty1_txt',
       'gname', 'guncertain1', 'individual', 'nperps', 'nperpcap', 'claimed',
       'weaptype1', 'weaptype1_txt', 'weapsubtype1', 'weapsubtype1_txt',
       'weapdetail',
       'nwoundte', 'property', 'ishostkid', 'dbsource', 'INT_LOG', 'INT_IDEO',
       'INT_MISC', 'INT_ANY'])
features = gtd_data[train_cols]
target = []
ans = features['success']
for val in ans==1:

    if val==True:
        target.append(int(True))
    else:
        target.append(int(False))



#gtd_data = gtd_data.replace(status_replace)
features = features.iloc[:, [1,2,3,4]]
features.dropna(axis='columns')
features = features.replace(0,np.NaN)
features = features.fillna(0)

# Fitting Decision Tree Regression to the dataset
from sklearn.tree import DecisionTreeRegressor
regressor = DecisionTreeRegressor(random_state = 0)
X_train, X_test, y_train, y_test = train_test_split(features, target, test_size = 0.3, random_state = 100)

#print(len(X_train)," = ",len(y_train))
regressor.fit(X_train, y_train)

# Predicting a new result
y_pred = regressor.predict(X_test)
#print("Confusion Matrix: ",confusion_matrix(y_pred.round(),y_test))
#print(type(y_pred.round()),type(y_test))
accuracy = accuracy_score( y_pred.round(),y_test)
print("Decision Tree Accuracy ",accuracy)
```

Figure 17: decision tree code

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules, and each leaf node represents the outcome.

In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

The decisions or the test are performed on the basis of features of the given dataset.

It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.

It is called a decision tree because similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.

In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.

A decision tree simply asks a question and based on the answer (Yes/No), it further split the tree into subtrees.

## 5. SVM (Support Vector Machines)

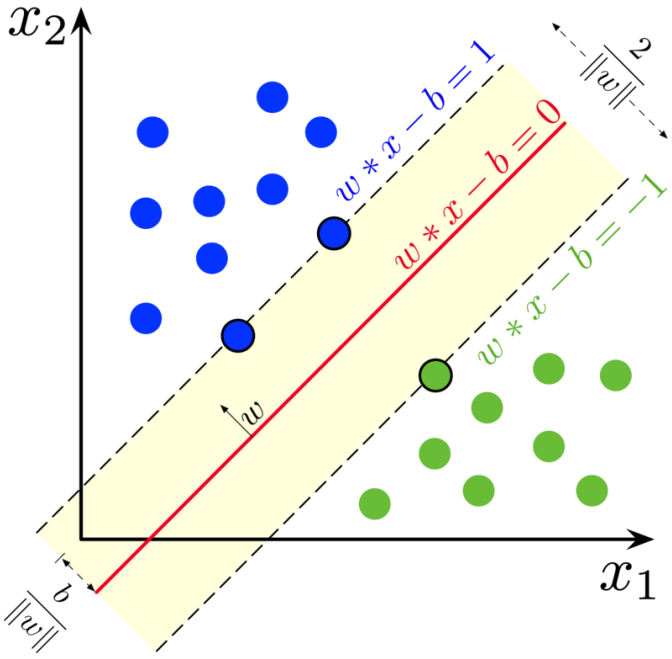



Figure 18: svm diagram

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

```
from sklearn import svm
regressor = svm.SVC()
X_train, X_test, y_train, y_test = train_test_split(features, target, test_size = 0.3, random_state = 5)
regressor.fit(X_train, y_train)
y_pred = regressor.predict(X_test)
#print("Confusion Matrix: ",confusion_matrix(y_pred.round(),y_test))
accuracy = accuracy_score( y_pred.round(),y_test)
print("Support Vector Machine Accuracy ",accuracy)
```

Figure 19:svm code

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane

# 6.TESTING AND VALIDATION

## 6.1. SYSTEM TEST

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

## 6.2. TYPES OF TESTS

### 6.2.1. UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is structural testing that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at a component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

### 6.2.2. INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event-driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfied, as shown by successful unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

### 6.2.3. FUNCTIONAL TEST

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input: identified classes of valid input must be accepted. Invalid

Input: identified classes of invalid input must be rejected.

Functions: identified functions must be exercised.

Output: identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

The organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows, data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified, and the effective value of current tests is determined.

## 6.3. SYSTEM TEST

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links, and integration points.

### 6.3.1 WHITE BOX TESTING

White Box Testing is a testing in which the software tester has knowledge of the inner workings, structure, and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black-box level.

### 6.3.2 BLACK BOX TESTING

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, like most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

### 6.4. UNIT TESTING

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct

phases.

Test strategy and approach

Field testing will be performed manually, and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages, and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

## 6.5. INTEGRATION TESTING

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

## 6.6. ACCEPTANCE TESTING

User Acceptance Testing is a critical phase of any project and requires significant participation by the end-user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

## 6.7. SYSTEM STUDY

### 6.7.1. FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase, and a business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis, the feasibility study of the

proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are,

- Economical Feasibility
- Technical Feasibility
- Social Feasibility

**1.ECONOMICAL FEASIBILITY**

This study is carried out to check the economic impact that the system will have on the organization. The amount of funds that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system was well within the budget, and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

**2.TECHNICAL FEASIBILITY**

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

**3.SOCIAL FEASIBILITY**

The aspect of the study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

# 7.RESULT ANALYSIS

## LOGISTIC REGRESSION CLASSIFICATION

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of the target or dependent variable is dichotomous, which means there would be only two possible classes.

PRECISION OF LOGISTIC REGRESSION:

```
In [45]: precision = float(false_positive)/float(true_positive + false_positive)
         precision

Out[45]: 0.8897395787432801

In [46]: # Data to plot
         labels = 'False Positive', 'True Positive'
         sizes = [1-precision, precision]
         colors = ['lightcoral', 'lightblue']
         # Plot
         plt.figure(figsize=(4,4))
         plt.pie(sizes, colors=colors, autopct='%1.2f%%', shadow=False, startangle=0)
         plt.title('Precision With Simple Logistic Regression', fontsize=12)
         plt.legend(labels, loc='lower left', fontsize=10)
         plt.axis('equal')
         plt.show()
```

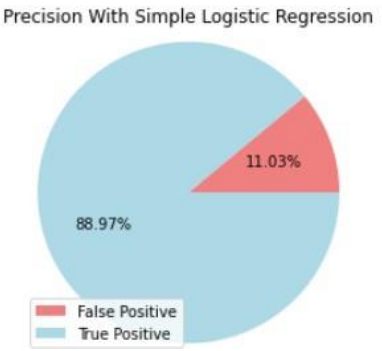



Figure 20: precision code of logistic regression:

ACCURACY OF LOGISTIC REGRESSION:

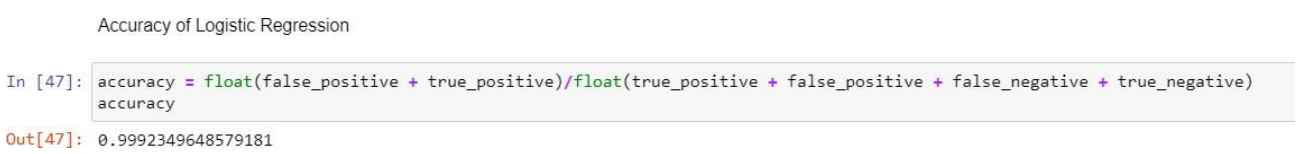



Figure 21: accuracy code of logistic regression

## K- NEAREST NEIGHBOURS:

k-NN is another algorithm commonly used for supervised classification problems. Each data points' k-closest neighbors are found by calculating Euclidean or Hamming distance and grouped into clusters. The k-closest data points are then analyzed to determine which class label is the most common among the set. The most common class is then classified to the data point being tested. For k-NN classification, an input is classified by a majority vote of its neighbors.

KNN ACCURACY:

```
X_train, X_test, y_train, y_test =  train_test_split(features, target,test_size = 0.25, random_state= 0)
sc_X = StandardScaler()
X_train = sc_X.fit_transform(X_train)
X_test = sc_X.transform(X_test)
classifier = KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p = 2)
classifier = classifier.fit(X_train,y_train)
y_pred = classifier.predict(X_test)
#check accuracy
accuracy = metrics.accuracy_score(y_test, y_pred)
print('KNN Accuracy: ',accuracy)
```

```
KNN Accuracy:  0.8864231776853136
```

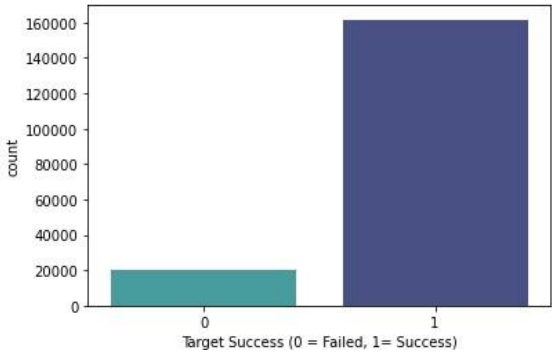



Figure 22: accuracy of knn algorithm diagram

## LINEAR DISCRIMINANT ANALYSIS:

ACCURACY OF LINEAR DISCRIMINANT ANALYSIS:

```
In [49]: from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
         from sklearn.tree import DecisionTreeClassifier
         lda = LinearDiscriminantAnalysis()
         regressor =  LinearDiscriminantAnalysis()
         X_train, X_test, y_train, y_test = train_test_split(features, target, test_size = 0.3, random_state = 100)
         regressor.fit(X_train, y_train)
         y_pred = regressor.predict(X_test)
         #print("Confusion Matrix: ",confusion_matrix(y_pred.round(),y_test))
         accuracy = metrics.accuracy_score( y_pred.round(),y_test)
         print("Linear Discriminant Analysis Accuracy ",accuracy)
```

```
Linear Discriminant Analysis Accuracy  0.8907132897923241
```

Figure 23: linear discriminant analysis diagram

**DECISION TREE:**

Decision trees classifiers apply questions and conditions in a tree structure. This approach applies decision rules inferred from the data features to predict the value of the target variable and create a model accordingly. The condition for categorization is included in the root and internal nodes. Inputs are entered at the top, and a tree is traversed down, following the branches. Once the input node reaches the terminal node, a class is assigned.

ACCURACY OF DECISION TREE:

```
# Predicting a new result
y_pred = regressor.predict(X_test)
#print("Confusion Matrix: ",confusion_matrix(y_pred.round(),y_test))
#print(type(y_pred.round()),type(y_test))
accuracy = accuracy_score( y_pred.round(),y_test)
print("Decision Tree Accuracy ",accuracy)
```

```
Decision Tree Accuracy  0.8895208042856094
```

Figure 24: accuracy of decision tree

**SUPPORT VECTOR MACHINES:**

Support Vector Machines: In machine learning, they basically come under the category of supervised learning, which analyzes data used for classification and regression analysis. An SVM model is a representation of points in space, mapped properly so that the categories get divided by a wide gap. If new examples are mapped, then they fall accordingly into the right side of the gap.

ACCURACY OF SVM:

```
In [51]: from sklearn import svm
         regressor = svm.SVC()
         X_train, X_test, y_train, y_test = train_test_split(features, target, test_size = 0.3, random_state = 5)
         regressor.fit(X_train, y_train)
         y_pred = regressor.predict(X_test)
         #print("Confusion Matrix: ",confusion_matrix(y_pred.round(),y_test))
         accuracy = accuracy_score( y_pred.round(),y_test)
         print("Support Vector Machine Accuracy ",accuracy)
```

```
Support Vector Machine Accuracy  0.8899060688339326
```

Figure 25: accuracy of SVM diagram

# 8.CONCLUSION

## 8.1. PROJECT CONCLUSION

The Machine learning techniques can help governments and law enforcement agencies to understand the factors of terrorism and to design strategies to deal with terrorism before a terrorist activity can actually happen

After training our models on the variable's month, Target_type, attack_type to predict the region of attack and country of attack it is estimated that Logistic regression, LDA, and SVM gives higher accuracy in both the cases on predicting Region and country of a terrorist attack. The results of the presented work can be used for enhancing defense against terrorist attacks in coming times.

Terrorism has become a huge threat to the world. A various Machine learning system, artificial intelligence and Using Global Terrorism Database (GTD) and Machine Learning Algorithms to Predict Terrorism and Threat Data-Analytics have provided us with a system to help the investigator and anti-terrorist or counter-terrorist squad to rapidly decide the most probable perpetrator responsible for a particular terrorist attack

## 8.2. FUTURE ENHANCEMENT

Machine learning algorithms have been used recently to study the different factors of terrorism popularity mainly because of the fact that a huge amount of labeled data is available in recent times.

The advancements in computer technologies have been able to create much powerful computer systems toper form the required computation in algorithms.

Models are used to make predictions of different factors that lead to terrorist activities.

The model is helpful for law enforcement agencies to make a prediction before an incident actually happens and potentially causes the loss of precious lives.

The predicted factors are explained below.

(i) Suicide: to predict whether a terrorist activity is going to be suicide or not.

(ii) Weapon type: to make a classification of the general type of weapons used in terrorist activity.

(iii) Region: to classify the region that will be targeted by the terrorist activity.

(iv) Attack type: to classify the type of attack carried out as a terrorist activity.

# 9.REFERENCES

[1] S. Sayad. Naïve Bayesian, from Predicting the Future. [Online],

[2] E.Fix and J.Hodges. Discriminatory analysis: nonparametric discrimination: Consistency properties. PsycEXTRA Dataset, (1951)

[3] https://cogsci.yale.edu/sites/default/files/files/Thesis2018Peng.pdf

[4] Mohammed, D. Y., & Karabatak, M. (2018, March). Terrorist attacks in Turkey: An evaluate of terrorist acts that occurred in 2016. In 2018 6th International Symposium on Digital Forensic and Security (ISDFS) (pp.1-3).

[5] Bang, J., Basuchoudhary, A., David, J., & Mitra, A. (2018). Predicting terrorism: a machine learning approach.

[6] Mathews, T., & Sanders, S. (2019). Strategic and experimental analyses of conflict and terrorism. Public Choice, 179(3-4), 169-174.

[7] Ravenscroft, I. (2019). Terrorism, religion and self-control: An unexpected connection between conservative religious commitment and terrorist efficacy. Terrorism and Political Violence, 1-16

[8] Khalifa, N. E. M., Taha, M. H. N., Taha, S. H. N., & Hassanien, A.E. (2019, March). Statistical Insights and Association Mining for Terrorist Attacks in Egypt. In International Conference on Advanced Machine Learning Technologies and Applications (pp. 291-300). Springer, Cham

[9] Wheatley, W., Robbins, J., Hunter, L. Y., & Ginn, M. H. (2019). Terrorism's effect on Europe's centre-and far-right parties. European Political Science, 1-22.

[10] Klenka, M. (2019). Major incidents that shaped aviation security. Journal of Transportation Security, 1-18.

[11] Guo, W., Gleditsch, K., & Wilson, A. (2018). Retool AI to forecast and limit wars.

[12] Hao, M., Jiang, D., Ding, F., Fu, J., & Chen, S. (2019). Simulating Spatio-Temporal Patterns of Terrorism Incidents on the Indochina Peninsula with GIS and the Random Forest Method. ISPRS International Journal of Geo-Information, 8(3), 13