*PIMA Indians Diabetes Database*

[*https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes*](https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes)

1. APA Citation: Lichman, M. (2013). *UCI Machine Learning Repository* [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
2. The terms of use can be found at <https://archive.ics.uci.edu/ml/index.html>
3. This data belong to females who are at least 21 years of age and of Pima Indian heritage. It contains their attributes like BMI (body mass index), Triceps skin fold thickness (mm), diabetes pedigree function (the effect of heredity on diabetes), their age and the number of times they have been pregnant.

This can be used to find correlation among the above mentioned parameters with the individual’s blood sugar levels or the individual’s susceptibility to diabetes – the extent to which these attributes either conform or negate the popularly held beliefs about their relation with diabetes.

1. This data can be used for training machine learning algorithms as well as to forecast the onset of diabetes mellitus as is mentioned in the following paper: Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., & Johannes, R.S. (1988). *Using the ADAP learning algorithm to forecast the onset of diabetes mellitus*. In Proceedings of the Symposium on Computer Applications and Medical Care} (pp. 261--265). IEEE Computer Society Press.
2. This data can be used to answer questions like:
   1. The factors and the extent to which they increase the susceptibility of a person in acquiring diabetes: a higher BMI or a family history of diabetes?
   2. The effect of multiple pregnancies on the susceptibility of developing diabetes – does it help to have kids at an earlier age rather than at a later age?
   3. The relationship between sub-cutaneous fat (Triceps skin fold thickness) and the tendency for developing diabetes.

*Global Terrorism Database*

<http://www.start.umd.edu/gtd/contact/>

1. APA Citation: National Consortium for the Study of Terrorism and Responses to Terrorism (START). (2016). *Global Terrorism Database* [gtd\_92to11\_0616dist (1)]. Retrieved from <https://www.start.umd.edu/gtd>.
2. The terms of use can be found at: <http://www.start.umd.edu/gtd/terms-of-use/>
3. This data has been developed by the University of Maryland in collaboration with the FBI and other similar organizations present globally and incorporates terrorism activities and relevant data from around the world. This database has been used for detecting areas that are susceptible to terrorist attacks and is updated at a regular interval. It could be quite interesting to step into the shoes of some of the world’s top most counter-terrorist agencies and to figure out how to use this data to make the world a safer place.
4. This data is used by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) in its effort to increase understanding of terrorist violence so that it can be more readily studied and defeated. Potential use may also include training machine learning algorithms to predict future terrorist activities as well as suggest typical measures to counter different categories/types of terrorist activities. It is widely used by many other government agencies world-wide.
5. Questions that can be asked based on this database may include :
   1. Whether it is possible to identify ‘*terror hot-spots’* on a map so that ample counter-terrorism measures could be deployed in the respective area?
   2. Would it be possible to predict and prevent future terror attacks based on this data?
   3. What kinds of counter-terrorism measures have been successfully deployed in the past to counter a specific kind of terrorism activity?
   4. Since this data is open to the public, and to the terrorist organizations too, would it affect their functioning?

*Flags*

<https://archive.ics.uci.edu/ml/datasets/Flags>

1. APA citation: Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
2. The terms of use can be found at <https://archive.ics.uci.edu/ml/index.html>
3. This data contains the names of 194 countries and the description of their national flags. The description of the flags includes the attributes like the number of vertical bars, the number of horizontal stripes, the presence of animate/ inanimate objects on the flags and other attributes amounting to a total of 30. This data can be particularly interesting since it delves on the question of whether it is possible to predict the dominant community/religion in a country based on its flag design.
4. This data can be used for training machine learning algorithms as well as for training artificial neural networks (ANN) as mentioned in the paper Wl/odzisl/aw Duch and Karol Grudzi nski and Grzegorz Stawski. *Symbolic features in neural networks*. Department of Computer Methods, Nicolaus Copernicus University.
5. Questions based on this data may include :
   1. Is it possible to predict the dominant religion/community in a country based on the design of its national flag?
   2. What can be said about the degree of inclusivity in a country or the prevailing form of the government – democratic or authoritarian by analyzing the attributes in its flag?
   3. What can be said about the effect of the zone/hemisphere in which the country is located, the size of its landmass and the population on the attributes of the flag?