

Jeremy Beard

20220829

MSDS 640

Week 1

ACM Code of Ethics and Professional Conduct

Introduction:

The Association for Computing Machinery (ACM) Code of Ethics (ACM Code of Ethics and Professional Conduct, n.d.) details the rules that each data scientist (and each computing professional in general) must follow to properly adhere to the standards that are required of data scientists (and computing professionals). Data can help influence major decisions made around the world and data scientists must be aware of that. They must act responsibly when confronted with large amounts of data which may be biased or even incorrect. Violations can easily occur if care is not taken to properly analyze and do the due diligence which is necessary to remain ethical when analyzing data.

The seven (7) general ethical principles from the ACM Code of Ethics are as follows:

1. Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing.
2. Avoid harm.
3. Be honest and trustworthy.
4. Be fair and take action to reduce discrimination.
5. Respect the work required to produce new ideas, inventions, creative works, and computing artifacts.
6. Respect privacy.
7. Honor confidentiality.

These rules are a code that data scientists must especially follow as they deal with large amounts of data which may influence important decisions made by individuals and/or corporations.

The ethical principles, and a description and example of each, are listed below:

General Ethical Principles

- 1) Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing.

This ethical principle firms the belief that data scientists need to seek to benefit society as a whole. It addresses ethical issues primarily as it deals with rules which govern the conduct of a person. Contributing to society and human well-being is simply promoting the quality of life of all people in society. An example of contributing to society would be promoting human rights and acting in such a way to protect human rights and autonomy.

2) Avoid harm.

Avoiding harming means just what it says: to mitigate all negative consequences of one's actions. Avoiding harm addresses social justice issues mostly as it seeks to provide positive impacts to society as a whole. It also addresses ethical issues as it provides a rule to live by (side note: all of these items in the ACM Code of Ethics provide a rule to live by). An example of avoiding harm would be to considering the total impact of one's actions before those actions are taken.

3) Be honest and trustworthy.

Being trustworthy means being honest, and usually the converse of that statement is true as well. Data scientists should be open and transparent about all actions and processes he or she used to analyze the data and reach the conclusions they reached. Making false or dishonest claims about data is a complete violation of the ACM Code of Ethics. Honesty and trustworthiness address ethical issues primarily as being honest is the first step in being an ethical data scientist. An example of being trustworthy and honest is an employee reporting a conclusion even though it isn't the conclusion anticipated or desired by upper management. This takes courage and a high level of respect for the truth and honesty.

4) Be fair and take action to reduce discrimination.

Being fair and not discriminating means to respect the equality of all human beings and having tolerance for others. This applies to data scientist because there is always an inherent bias in a dataset. A data scientist should be aware of this and should work to reduce bias as much as possible. This will lead to the most truthful conclusions and will ultimately lead to the best decisions. Fairness and equality address social justice issues primarily as they seek to promote fairness to the society as a whole. An example of being fair and not discriminating is to take care to have a sample representative of the population in one's data. If a sample is not representative of the actual population, then the conclusions gleaned by the analysis are not fair relative to the population. There will be an increased level of bias in the analysis if the sample is not actually giving a conclusion relevant to the population.

5) Respect the work required to produce new ideas, inventions, creative works, and computing artifacts.

Respecting the work required to produce new concepts means respecting the work another individual does to create something. Whether that be an analysis, a chart, a piece of art, we must take care not to steal others' work as they have put a significant effort into it. Respecting the work required to create something new means putting in the work to create something new yourself, rather than taking from your neighbor. It especially means not claiming ownership of publicly shared resources. Respecting others' work addresses ethical issues as it is highly unethical to steal others' work or claim it as your own. An example of this concept would be

analyzing a problem a different way than a neighbor, perhaps to reach a different conclusion but overall to broaden the scope of our understanding of the problem.

6) Respect privacy.

Respecting privacy, in the context of a data scientist, applies heavily to personal information and personal data. Data scientists should only use personal data for legitimate reasons and should not attempt to take advantage of others' personal information. Data scientists should minimize the amount of identifying personal information collected during an analysis or experiment. Respecting privacy obviously address privacy issues as a whole, more than social justice or ethical issues (although this can address ethical issues too; they all can). An example of this would be anonymizing the data to remove one's identity from their data gathered.

7) Honor confidentiality.

Honoring confidentiality is similar to respecting privacy, in that it helps work to limit the number of identifiers someone could use to link a datapoint to a specific individual. This helps keep the data anonymous while still being able to reach valid conclusions that are useful. Honoring confidentiality addresses privacy issues the most, while also addressing ethical issues as privacy can be considered an ethics issue. An example of honoring confidentiality would be to delete the identity column, or first/last name column, from a data table to remove all identifying information from the dataset.

Conclusion:

In conclusion, the seven (7) ACM Code of Ethics is a great list of rules to start from when referencing how to be ethical as a computing professional and especially as a data scientist. It details that we should protect one's personal information, not falsely claim ownership of work, reduce bias and increase fairness, be honest, avoid harm, and contribute to the common good. These are all positive actions to help increase the level of goodness a person contributes to the world, especially as a data scientist. This is also not a list to simply read one time and then forget about. This is a list to continue to re-visit as one progresses along his or her career. It should also grow and develop as one's understanding of the world grows and develops.

References:

- 1) ACM Code of Ethics and Professional Conduct. (n.d.). ACM. Retrieved August 28, 2022, from <https://www.acm.org/code-of-ethics>