Should You Open-Source Your Model?
Ethical questions for open-sourcing Machine Learning models

Robert Munro
Independent Researcher
San Francisco, CA
rmunro@alumni.stanford.edu

ABSTRACT
This paper covers a set of ethical guidelines to help data scientists decide whether to open-source their Machine Learning models. Many questions are existing ones related to open-source data while some are new questions that are specific to Machine Learning. Examples are drawn from real uses of open data and Machine Learning in disaster response and healthcare over the last decade, sharing lessons learned. A set of questions are provided that data scientists should ask when deciding whether or not to open-source a model. It is recommended that data scientists create a “Model Statement” to be explicit about a model’s capabilities and to help with the decision.

1. INTRODUCTION
In the last few years, Machine Learning models in Natural Language Processing (NLP) have been able to generalize their capabilities very easily for the first time. For example, Google’s BERT system [5] was announced in November 2018 and demonstrated state-of-the-art accuracy on many very different language tasks, including question-answering, sentiment analysis, and identifying named entities in text. Earlier the same year, the Allen Institute’s ELMo system [19] showed similar state-of-the-art results. ELMo and BERT won best paper awards at a top NLP conference, NAACL, in successive years as a result. There have been a large number of new models in the short time since, including from the same researchers, with newer state-of-the-art results. This was big step in Machine Learning. Previously, almost all models in Natural Language Processing could only perform one task. For example, you would have one model for question-answering, a second model for sentiment analysis, a third model for identifying entities, etc. While people had built multi-task NLP systems in the past, there had never been models that were broadly multitask and state-of-the-art. The main take-away from ELMo, BERT and related efforts is that they can be adapted to almost any NLP task in the languages and domains that they are trained in. They are known as pre-trained models: they are pre-trained on very large amounts of data and can then be easily tuned to new tasks.

For the first time in history, therefore, we don’t know what our most powerful Machine Learning models are capable of.

We haven’t come close to testing all possible use cases for pre-trained models, and all evidence points to the fact that the models can become more powerful if they have more data. This has raised some concerns about misuse. Machine Learning researchers have always been encouraged to open-source their code, models, and data. But some researchers have questioned whether we should open-source models without question because of the potential negative use cases.

In Computer Vision, we are seeing similar advances in creating videos that seem real but are not, known as Deep Fakes [21]. While Computer Vision has had advances in Transfer Learning for a few years longer than NLP and the Deep Fakes technology is particularly compelling, Computer Vision does not (yet) have pre-trained models with the same adaptive power for multitask learning. So, this paper covers both NLP and Computer Vision, but focuses on the former.

2. WHAT CAN WE LEARN FROM PAST EXPERIENCES?
Machine Learning is a fast-moving field that deserves more attention in ethics. It has been five years since a KDD Workshop on Data Ethics [7] and the ethical issues raised there still remain. We have a clearer idea on the response needed for some of the ethical issues raised five years ago, while newer ethical questions have been raised since that time due to more recent Machine Learning advances, like the power of the new pre-trained NLP models.

This is a position paper that draws on experience in Machine Learning for disaster response and healthcare. I worked in post-conflict development in Sierra Leone and Liberia before I moved to US to complete a PhD in Natural Language Processing at Stanford, and have worked in Machine Learning applied to healthcare [14; 8] and disaster response [9; 15] since then. However, only a handful of the Machine Learning papers published included discussion of ethics [13; 16; 10; 17] and the papers focused purely on ethics were in non-academic formats, like invited talks or blog posts [11; 12]. That meant that Machine Learning researchers missed much of the ethical discussion that went alongside this work.

So, this paper focuses on best practices about the ethics of open data and deploying Machine Learning models, to complement the algorithm-focused papers that were previously published about the same events. It is fair to say that most people in the Machine Learning community have rarely needed to consider the sensitivity of their data and models...
pre-trained models are not a step in that direction. There are plenty of things that Intelligence (AI that acts autonomously) or necessarily even the first time, that is not the same thing as General Artificial exactly what our most powerful models are capable of for This paper also addresses the hype. While we don’t know before, so this paper aims to bridge that gap.

While the focus of this paper is on the potential negative use cases, it also covers the issue of representation. For example, the implicit lack of representation in a model that is trained only on English data means that most of the world will not get to take advantage of its positive use cases. The model will therefore amplify existing biases towards speakers of privileged languages like English.

2.1 How is an open model similar to open data?

An open model is similar to open data, in that data can be re-purposed for many tasks. In fact, the pre-trained models like ELMo and BERT can be thought of as generalizations over the data. The most widely used pre-models today are built by predicting the order of words in sentences and the order of sentences themselves. So, the data/model analogy has technical grounding. For example, with any large dataset, someone could construct a single-purpose system that created fake news, or they could create a system by tuning a pre-trained model on the that same data to create fake news. Similarly, someone could create a single-purpose system to extract Personally Identifying Information (PII) from text, or they could tune a model pre-trained on that same data to identify PII. So, the same ethical considerations that we apply to open source data can also be applied to models built on open data.

2.2 How is an open model different to open data?

Open data and models built on open data differ in two ways: it is easier to tune a pre-trained model than to build a single-purpose model from scratch, and there is more data security in releasing a model than in releasing the data that the model was trained on.

1. The ease of tuning a model

It can take orders of magnitude more resources to build a system from raw data, compared to building that system by adapting a pre-trained model. In security, you treat every strategy as “breakable” [20]. The goal is to make the cost of breaking some security measure higher than the value of the data that you are protecting. Therefore, it is not worth a bad actor’s time to break the security. The same security practice can be applied here: how much easier is it for bad actors to use the model maliciously? For example it has always been possible for bad actors to create fake news. However, it might have taken a very large number of Machine Learning scientists to create a system to do this, with much of it still a manual process. This is a scale of effort that might be off-limits to anyone but large (possibly state-sponsored) actors. Today, a lone internet troll has the ability to create a large scale fake news system.

2. Better (but not perfect) security in a model

An open model is also different from open data in that it can better hide sensitive data. This is a grey area, as it is possible to reverse engineer Machine Learning models to reveal the data they have trained on. Privacy-preserving Machine Learning is an established field [6], but it is very much an unsolved problem. We can apply the same security question here: is the cost of reverse engineering the model greater than the value of the information in the original training data?

One important difference for sensitive data is how the discovery of sensitive information can be treated. Open models have an advantage over open data because you can safely re-release the same model with sensitive data removed. You cannot do the same with open data. For example, imagine that you release a data set and then discover that the data has sensitive information like someone’s personal address. You can’t simply remove that address and re-release the data, because it will be trivial for someone to look at the difference between the data set before/after, and use the difference to discover where there was sensitive information. So, re-releasing open data with sensitive data scrubbed makes it worse: you are basically labeling where the sensitive data occurs for any bad actor to take advantage of! By comparison, there is no similarly trivial way to see the difference between two models. If the first model was able to generate someone’s real address, that address information was spread across many nodes in the model. The distribution of information across nodes in the new model will be completely different from the old model, so there will be no simple difference between the models that could locate sensitive information in the first model that was scrambled from the second.¹

3. QUESTIONS TO ASK WHEN THINKING ABOUT OPEN-SOURCING A MODEL

This section highlights existing examples of best practices and lessons learned from past uses of open data or Machine Learning in contexts including: man-made disasters like wars and civil unrest; natural disasters that make people vulnerable like disease outbreaks, earthquake and storms; healthcare; and election monitoring.

The following questions are addressed:

1. Should I open-source my model, just because everybody else in my field does?

2. My model is 100% from public data, so do I need to question whether or not to open source my model?

3. How do I evaluate risk?

4. Should I trust journalists to make decisions about risk?

5. How do I ensure that journalists are talking responsibly about my Machine Learning research?

¹This might be partially possible with GANs, but I am not aware of research into comparing models with/without data scrubbed in order to identify sensitive information. Any solution would be harder to implement than simply finding the diff of two data files, and much less accurate.
6. Should I trust governments to make decisions about risk?
7. Should I investigate solutions to the negative use cases from my model?
8. Is this a new problem in Machine Learning?
9. Is fake news a new problem?
10. Should I focus on balancing the bad use cases for Machine Learning with ones that are more clearly good?
11. How can I be explicit about the limitations of my model?

For each question, the answers draw on best practices with examples and pointers to existing discussion in venues that data scientists might not typically read.

The answers to each question of these questions will also draw on recent work within the Machine Learning community, like Bender & Friedman’s work on Data Statements to mitigate bias in Natural Language Processing [2] (presented at the same conference as BERT) and the AI Now Institute’s recent report on bias and representation in AI [18].

3.1 Should I question whether or not to open source my model?
Yes. If your model is built on private data, it could be reverse engineered to extract that private data.

3.2 My model is 100% from public data, so do I need to question whether or not to open-source my model?
Yes. Non-sensitive data can become sensitive if you republish it in new contexts, and aggregated data can become more sensitive than its individual data points. We can include Machine Learning models as a form of “aggregated data” in this context.

During the Arab Spring, I saw a lot of people tweeting about their local conditions: road closures, refugees, etc. While they were “public” tweets, the tweets were clearly written with only a handful of followers in mind, and many people did not realize that reporting road closures would also help paint a picture of troop movements. As an example of what not to do, some of these tweets were copied to UN-controlled websites and re-published, with no mechanism for the original authors to remove them from the UN sites. Many actors within the Middle East and North Africa saw the UN as a negative foreign influence (or invader) and the people tweeting were therefore seen as collaborators: the bad actors didn’t care if these people only intended to be sharing information with a small number of followers.

State-sponsored actors repeatedly tried to access sensitive information in this and related humanitarian efforts [3], and we know they had some success discovering personal information about people who were reporting local military movements.

So, you need to ask yourself: what is the effect of recontextualizing the data or model so that it is now published by myself or my organization?

It is also very common that aggregated data is sensitive when individual data points are not sensitive. This is standard practice for a lot of military organizations: when they aggregate data from one set of sources, they reevaluate that aggregated information for its level of sensitivity. The aggregations are typically the results of statistics or unsupervised machine learning, but a supervised model built on that data would equally apply. For a public example of how aggregation changed the sensitivity of data in the military, see the recent case where runners in the military using Strava accidentally gave away the location of bases, when it showed heat-maps of where people in the military ran the most.²

Many organizations have chosen to adopt similar policies. For example the publishing organization, Medium, includes this restriction (my emphasis) “dox [expose the identity] someone including by exposing personal information or aggregating of public information”.³

So, even if your model is built on open data you should always ask yourself: is the aggregation of the data in my model more sensitive than the individual data points?

3.3 How do I evaluate risk?

Use the same strategy as the security industry to weigh up risk: weigh up the cost of misuse compared to the perceived value it provides for a bad actor.

In security, you treat every strategy as “breakable”. The goal is to make the cost of breaking some security measure higher than the value of the data that you are protecting. Therefore, it is not worth the bad actor’s time to break it. Is the cost to reproduce the models from your research papers worth the effort for someone who wants to use them for negative reasons? You should be explicit about this. This is then one factor that will go into your decision about whether to open source or not. (See more below on how to express this in “Model Statements”.)

3.4 Should I trust journalists to make decisions about risk?

No. Or at least, not without question. Journalists have to sell content and the more sensational content tends to sell more. Consciously or unconsciously, a journalist might lean towards open-sourcing, as it is then easier for them to write about it. On the other hand, a decision to not open-source the data could lead to a sensationalist report about the dangers that motivated that decision. Either way, your organizational goals are not aligned with those of journalists to the point that you can blindly trust them.

I’ve seen many cases of preventable deaths caused by bad journalism in disasters. In the 2014 Ebola outbreak in West Africa, I predicted:

“For every person who contracts Ebola, ten people will die nearby from other preventable conditions. The silent victims of Ebola.”⁴

I was an advisor to most major governments and aid organizations responding to the outbreak because I had lived in the region and had separately worked in epidemic tracking for Ebola in East Africa. I warned almost every news

⁴https://medium.com/@robert.munro/the-silent-victims-of-ebola-e1a8854185a9
organization, both local and international, about the hype around reporting and the deaths it could lead to.

A small group of press listened, but most did not. When the deputy head of Sierra Leone’s Health Ministry spoke at a conference in San Francisco following the end of the epidemic, she reported these exact same sad figures: they estimated that ten people died from other diseases because they were avoiding clinics, for every one person that died directly from Ebola.

So, be very careful when talking to journalists about your research.

3.5 How do I ensure that journalists are responsibly talking about my Machine Learning research?

Get media training. Many organizations run media training and even just a few hours will help. I can’t summarize everything you need to know in this paper, but here’s the most important thing I have found that generally works: 

Ask the journalist what their story is about.

If they are writing about advances in Machine Learning research, then you are probably ok. If they are writing about “Dangers in AI”, or “Fake News”, or “Interference in Elections”, then you should be increasingly careful about how your interview might be skewed to fit their narrative.

The strategy of asking about what a story is about has worked for me with only one exception that I remember: a BBC reporter was writing an article about how English was dominating the internet. I gave an interview saying that, no, English’s share of the internet has been steadily declining; people prefer their primary language and English is becoming a “second language of the internet”. But they reported me as saying that “English is becoming the language of the internet”. If this happens, there’s not much you can do: the BBC had broader reach than my tweet denying I said this. You can ask the media organization to amend the article or at least make a public statement that you were misquoted.

So, understand that there is no guaranteed solution to your voice being the loudest about your own model if the story is picked up by the media, and use caution when talking about your model as a result.

3.6 Should I trust governments to make decisions about risk?

No. Obviously, you shouldn’t break the law. But just because it is legal, that doesn’t mean it is ok. Governments are a group of people like any other, trying to get their heads around the real and not-so-real threats of Machine Learning.

Governments are also prone to media influence. The Liberian government shut its borders in response to the Ebola outbreak mentioned earlier. Having worked on those borders, I know this was a complete farce. The borders are a series of connected villages, rivers, creeks, and forest trails that existed long before the official borders today. “Closing” the borders and sharing data about the infections along the borders made the government seem decisive, but ultimately would have protected almost no-one from Ebola. Instead, it drove people away from clinics for other treatable illnesses, making the situation much worse.

Like journalists, treat governments as important potential partners, but realize that you all have different agendas and many of them will be aligned, but not all of them.

3.7 Should I investigate solutions to the negative use cases from my model?

Yes! For example, if a model can create fake news, can you also create a model that can distinguish between real and fake news? Researchers including many of ELMo’s creators did exactly this [22]. If you can demonstrate, quantitatively, that a negative use case for the data is easier or harder to combat, then that will be one factor in your decision-making process for whether to release a model openly or not.

3.8 Is this a new problem in Machine Learning?

No, and you can learn a lot from past experience.

In 2014-2015, I was approached by the Saudi Arabian government on three separate occasions to help them monitor social media for dissidents. At the time I was CEO of Idibon, a company who had the most accurate Natural Language Processing technology in a large number of languages, so we were naturally seen as technology that could be the best for their use case. We were first approached directly by a Saudi Arabian ministry, and then indirectly, once through a boutique consulting company and then once again through one of the five biggest consulting companies in the world. In every case, the stated goal was to help the people complaining about the government. After careful consultation with experts on Saudi Arabia and Machine Learning, we decided that a system that identified complaints would be used to identify dissidents. As Saudi Arabia is a country that persecutes dissidents without trial, often violently, we declined to help.

If you are facing a similar dilemma, look for people who have the depth of knowledge to talk about the community who would be most affected (ideally people from within that community) and people who have faced similar Machine Learning problems in the past.

3.9 Is fake news a new problem?

No. Propaganda is probably as old as language itself.

In 2007, I was escorting journalists reporting on the elections in Sierra Leone when we kept hearing reports of violence. In 2014-2015, I was approached by the Saudi Arabian government to help them monitor social media for dissidents. At the time I was CEO of Idibon, a company who had the most accurate Natural Language Processing technology in a large number of languages, so we were naturally seen as technology that could be the best for their use case. We were first approached directly by a Saudi Arabian ministry, and then indirectly, once through a boutique consulting company and then once again through one of the five biggest consulting companies in the world. In every case, the stated goal was to help the people complaining about the government. After careful consultation with experts on Saudi Arabia and Machine Learning, we decided that a system that identified complaints would be used to identify dissidents. As Saudi Arabia is a country that persecutes dissidents without trial, often violently, we declined to help.

If you are facing a similar dilemma, look for people who have the depth of knowledge to talk about the community who would be most affected (ideally people from within that community) and people who have faced similar Machine Learning problems in the past.
3.10 Should I focus on balancing the bad use cases for Machine Learning with ones that are more clearly good?

Yes. It is easy to have a positive impact on the world by releasing models that have mostly positive application areas. By contrast, it is difficult to have a positive impact on the world if your model has many negative application areas, even if you decide to then keep it private.

This is a failing of many recent pre-trained models because they were released only on English data. English only makes up 5% of the world’s conversations daily. English is an outlier in how strict the word order in sentences needs to be, in standardized spellings, and in how ‘words’ are useful as atomic units for Machine Learning features. Most pre-trained NLP models rely on all three: word order, words as features, consistent spellings. Would the same methods even work for the majority of the world’s languages? We don’t yet know.

If you don’t want to step into the grey area of applications like fake news, then pick a research area that can have an impact more directly, like language models for health-related text in low resource languages.

3.11 How can I be explicit about the limitations of my model?

If you are making a model open, you should be explicit about the capabilities, limitations and known vulnerabilities. For some time, it has been standard for researchers releasing language data to include “Data Statements” which list the content of the data including any limitations. This is especially true for linguists working in language documentation for less widely-spoken languages, as text from or about marginalized communities can be the most sensitive [1]. Bender & Friedman recently presented a position paper recommending “Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science” [2], arguing that all papers in Natural Language Processing should include Data Statements.

We can extend the Data Statement concept to Machine Learning models, as recommended by this paper and covered in the next section.

4. MODEL STATEMENTS

A model statement is a characterization of a model that provides context about how the model might generalize, how the software associated with the model might be deployed, what biases might exist in the model, what known exploitations exist in the model, and ways to mitigate those exploitations (if they exist). It is super-set of the data statements in Bender & Friedman [2], following their invitation to extend their suggestions for Data Statements for NLP. The additions in this paper focus on the issues associated with models that are beyond those associated with data, and are generalized across different Machine Learning use cases, not just language data.

It is not the intention of a Model Statement to be entirely negative. An important component of talking about the negative cases is their limitations. It may be obvious to a data scientist that a pre-trained NLP model cannot be taught to fly a killer drone. However, someone outside of the field might not be able to distinguish seemingly autonomous behavior in natural language from autonomous behavior controlling robots. Model Statements are therefore a chance to get ahead of the erroneous conclusions that someone might make.

A long form model statement definition is given below. In addition to the long form statement, a short form summary statement might be only a couple of sentences long, summarizing the model and potential uses and limitations. A short-form is useful to provide in an abstract or README file for a model, to give the reader context right away, so they are reading about the model in an informed way up-front. A longer data statement can accompany the paper or code as an appendix or as a separate file in a code repository.

4.1 Long Form Model Statement

A. Curation Rationale. What data sources were included and what were the goals in selecting that data? From what sources and why only for those sources?

B. Location and Languages. What locations did the data come from? You should provide an estimate if you don’t know for certain. For example, if it is images from a certain online collection and 75% of the people posting to that collection are in the USA, then you should state the 75% of the images are most likely from the USA and also provide the breakdown (if known) of other locations. For written, spoken, and signed data, what ISO language code(s) are in the data? Are there dialectal or location-based limitations? (eg: Southern UK English only).

C. Demographics of the data creators. What is the breakdown of the data sources by age, race/ethnicity, gender, languages spoken, socio-economic status, and other demographics that might impact the model performance? As with location, you might need to make an estimate from where the data is taken. For example, if you built a pre-trained model on Wikipedia’s English data, you should state that as much as 90% of the data was created by men, as that is the most accurate estimate of Wikipedia’s editors.5

D. Annotator Demographics. What are the locations and demographics of people who annotated the data? Include annotation guidelines. Note if any part of the guidelines rely on judgments that could result in annotations that correlate with the demographics of the annotators. The correlations could be biases, or errors arising from familiarity. For example, you might have annotators from only one part of the world and you asked them to label images of types of clothing from a different part of the world, some of which they will not be familiar with.

E. Context of Data Creation. For language, people will speak/write/sign very differently when they are communicating formally than when they are chatting with friends and family. For images, photos taken with intention of being posted on social media will be chosen and framed very differently from photos taken at 90% of Wikipedia’s Editors Are Male—Here’s What They’re Doing About It. https://www.theatlantic.com/technology/archive/2013/10/90-of-wikipedias-editors-are-male-heres-what-theyre-doing-about-it/280882/
random intervals from an autonomous vehicle. Understanding the time, place, modality, and intended consumer of the data will therefore help understand the capabilities of models built from that data.

F. Domain. Is the data from any domain, genre, register or topic that is relevant to how the model will perform? Examples include “Book Review”, “Fashion Photography”, etc. Models tend to be very domain-specific in their application areas, so if the data is restricted to a particular domain, then the model will be, too. You might want to think carefully about where multiple domains exist in a source. For example, the “political news” domain will have news articles that include quotes from politicians. So, there is also a second “political announcement” domain within some documents, which means that one negative use case could be generating fake announcements from politicians that sound genuine.

G. Data Quality. Especially for audio, image, and videos, this includes the quality of the camera/recording equipment, the resolution, the type of compression used, or any other technical aspect of the data that could influence the model. Include everything that could change the data, from the original creation of the data up until you trained on it. For Computer Vision models, if you have down-sized or distorted the data for your model (eg: converting every image to 256x256 pixels and ignored the transparency channel) then include that information here. For NLP models, if you have lower-cased the text or ignored all characters from outside of some Unicode range, include that here. The model’s performance will often be limited by the quality, filtering, and data processing decisions that were made before the model was built.

H. Primary Functions. The functionality of the model as it is released, without further tuning. For example, you might have trained a model on ImageNet and have it return the top-5 predictions, so the model function is that it takes an image and returns the top 5 most likely labels for that image. The expression “function” is used in the general sense here, but you can also think of “function” in the software sense: if your model was a function/method, what input parameters does it take and what is returned? For a pre-trained model, the function will describe how it was pre-trained. For example, for a bi-directional word-only model, you might state that your model predicts words (or word segments) in a sentence, given the surrounding words.

I. Derivative Functions. The functionality that model can be derived to perform and the resources required for each function. For example, if you have a general pre-trained model like BERT, you could make a statement about how the model can label whole sentences, label segments within a sentence, or generate novel sentences in the same languages and domains as the training data. You could then list specific examples, like labeling for sentiment, extracting named entities, summarizing text, etc. For each use case, you would state that the derivative functionality requires substantial additional human labeled-text. On the other hand, if the model can be adapted to new use cases with only a handful of examples, then this would be important to note. This will directly inform the evaluation that we have adopted from security: how easy it for someone to adapt the model for a bad use case, relative to the value they might get from the resulting model?

J. Functional Limitations. Some of the limitations will be implicit in the other items, but they should be spelled out because what is implicit to one person might not be to another. For example, if your model is trained only on English news articles and Wikipedia, then you should explicitly state that it can’t be adapted to other languages and probably not easily to other domains of text, like interactive English chat messages. Instead, someone would have to take the code and train a new model from scratch. This highlights potential bad use cases from similar models that aren’t possible with yours.

K. Mitigation Strategies. How can you limit or combat the bad use cases? If the model can be used to generate fake news, then you can include ways to stop this. For example, you could state that a different model could be used to determine if news was real or fake. The creators of ELMo and additional colleagues did exactly this [22]. List all the potential negative use cases and ways to avoid them or combat them. Be explicit if you are not aware of any ways to combat some potential bad use cases.

L. Other. Any other relevant information that could be important for the model’s performance and potential use cases.

M. Provenance Appendix. If the model is itself adapted from any prior model or if any of the data used has a prior Model Statement or Data Statement, then include those.

5. CONCLUSION

If you are thinking about open-sourcing your model, write a Model Statement first. Use this statement to inform your decision about whether to open-source the model, keep the model private, or do something in between like only release the model to researchers working under an Institutional Review Board (IRB).

If you do open-source the model, a Model Statement will help others make their own evaluation.

Acknowledgements

I am grateful to Emily Bender for allowing me to read early versions of the “Data Statements for NLP” paper that she and Batya Friedman [2]: it shaped a lot of my thinking. I also thank my co-panelists on “This Week in Machine Learning and AI” [4] for their discussion on open-sourcing models, hosted by Sam Charrington: Anima Anandkumar,
Amanda Askell, Miles Brundage, and Stephen Merity. Their discussions during and after the panel were deeply valuable. Thank you to everyone that I have worked with in health and disaster response.

6. REFERENCES


[7] K. Jacobs and R. Munro. (Organizers) KDD Workshop on Data Ethics. ACM Conference on Knowledge Discovery and Data Mining (KDD), 2014.


