SUPPORTING DOCUMENTS PARTICIPANTS

D4G CHALLENGE - 2020

TABLE OF CONTENTS

1. INTRODUCTION TO D4GC AND EMERGENT

In 2015, Emergent Leuven was founded by students from the Faculty of Engineering at the KU Leuven. Their mission was to create an organization built around multidisciplinarity. Today, Emergent Leuven is the **data science organization** at the KU Leuven.

There is a shortage of professionals who master an advanced set of skills in Data Science, but an increasing demand for these skills. With Emergent Leuven, we solve this problem in 3 ways:

1. Develop transdisciplinary analytical skills (data science + communication + strategy) through participation in workshops and educational tracks, as well as organizational skills through the organization of activities.
2. Apply skills at challenges and consultancy tracks.
3. Transfer mastered skills by teaching a workshop, or coaching during challenges or consultancy tracks.

We do this in close collaboration with industry and academia to bring students, researchers, and professionals closer together.

Our flagship event, the **Data4Good Challenge**, allows students to solve a real case. Participants work together in multidisciplinary teams. Together, they are tasked with solving a humanitarian problem through the use of data analysis and visualization. In a truly Emergent fashion, participants will have to think through all aspects of their solution, not only considering the economical effects but also the social and ethical consequences of their business plan. Since the challenge is two-fold - produce insights, then pitch these insights as a solid business case - an interdisciplinary collaboration is the best way to tackle this challenge successfully.

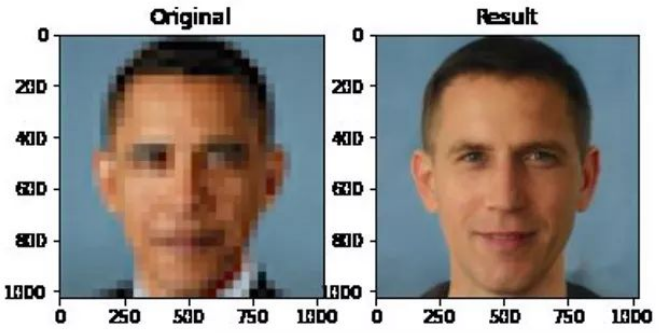
2. SCHEDULE

08h45	Welcome and Introduction to the Data4Good Challenge
08h50	Partner presentations
09h15	Introduction to the challenge
09h30	Start of the challenge
10h00	Start of the coaching sessions by our partners. See the individual schedule to see when the coaches will visit your table.
14h00	Submission deadline
15h00	Start of the pitching sessions. See the individual schedule to see when and where your team needs to pitch your case.
20h00	Start of the reception
20h10	Prize ceremony
20h30	Announcement winners tombola and virtual networking
...	

3. INTRODUCTION TO CHALLENGE

The problem of discrimination in machine learning has been described as one of the most important challenges in artificial intelligence by leading experts in the field. A recent twitter interaction between Yann LeCun, the father of deep learning and one of the most prominent data scientists in the world, and Timnit Gebru, co-founder of the Black in AI group and co-lead of the ethical AI team at google, received a lot of attention in the AI community.

The Twitter discussion was about a model created by Duke University that transforms low-resolution images into high-resolution images. The model transformed a low-resolution image of former US President Barack Obama into a picture of a white person. This led to an intense discussion about the bias in machine learning models.



The problem indicated above is only one of the many cases in which machine learning models showed questionable behavior against minority groups. The abundance of examples of discrimination in machine learning models has encouraged both academic and professional researchers to explore the topic of fairness in AI.

The **US justice system** makes use of algorithms that assign risk scores to offenders of various crimes. These scores are then used by judges to determine the type and length of the sentence. One such algorithm used by the states of New York, Wisconsin, California, Florida's Broward County, and other jurisdictions is the Compas algorithm developed by Equivant (formerly known as NorthPointe Inc.).

In 2014, U.S. Attorney General Eric Holder warned that the risk scores might be injecting bias into the courts. He called for the U.S. Sentencing Commission to study their use. **In the end, the commission did not study these algorithms.**

In recent years, more and more data about these algorithms became publicly available and now it is the perfect time to properly study these algorithms. **Your team's job is to study whether or not these algorithms cause bias in the US Justice System and how this affects various stakeholders.**

4. DISCUSSION OF DATASETS

To solve this year's challenge your team can make use of three different datasets. The datasets are **complementary to each other** which means that you need to combine insights from multiple datasets to properly tackle the challenge.

1. Compas Dataset

The Compas algorithm is one of the methods used in the US Justice System to determine the recidivism risk scores for defendants. This dataset contains background information about the offenders and scores that have been assigned to them by the Compas algorithm.

The Compas scores are calculated based on a questionnaire but the information gathered by this questionnaire is not publicly available. Nonetheless, this dataset can give insight into how the algorithm works.

The dataset also contains information about whether or not the defendant recommitted a crime in the two years following their release. This information can be used to study whether or not the algorithm is good at predicting recidivism and whether it does this in a fair way.

Variables

Case number	Unique identifier
First name	The first name of the offender
Last name	The last name of the offender
Sex	The gender of the participant (Male-Female)
Age	The age of the offender at the time of the assessment by the Compass algorithm
Age category	The age category to which the offender belonged at the time of the assessment by the Compass algorithm
Race	The race of the offender
Juvenile[1] felony[2] count	The number of juvenile felonies committed by the offender
Juvenile misdemeanors count	The number of juvenile misdemeanors committed by the offender

[1] Charges handled by the juvenile court (young offender's court)
[2] a crime regarded in the US and many other judicial systems as more serious than a misdemeanor.

Juvenile other count	The number of prior juvenile convictions that are not considered either felonies or misdemeanors
Priors count	The number of prior convictions. It does NOT equal the sum of juvenile convictions as it also includes non-juvenile convictions
Charge description	A text description of the charges against the offender
Charge degree	A measure for the severity of the charges against the offender
Offense date	The date at which the offense was committed
Screening date	The date at which the offender was screened by the Compas algorithm
Jail in	The date at which the offender entered jail
Jail out	The date at which the offender left jail
Type of assessment	The Compas algorithm computes various scores. We only provided information on the Risk of Recidivism assessment.
Score	**The score that is given by the algorithm ranging from 0 (low risk) to 10 (high risk)**
Score text	Score categories (Low, Medium, High), these categories have not been defined by the Compas algorithm and are subjective.
Is recidivist	**Shows whether or not the offender committed another crime in the two years following their release from prison**
Prediction	**Shows whether the algorithm predicts the accused of reoffending (1: will reoffend, 2: will not reoffend). Defined as score being greater than 6.**

2. US Department Of Justice summary statistics

The second dataset provided by the US department of justice provides summary statistics of the number of arrests for various crimes per race per year (2010-2018). This dataset can be used as a way to create additional context to the problem at hand.

Citation: OJJDP Statistical Briefing Book. Estimated number of arrests by offense and race, 2018.

3. Pre Trial Risk Dataset

The pretrial risk dataset is a collection of documents that contain information about all the algorithms that are currently used in the US Justice system. The dataset can be most easily accessed using excel. There are six different sections:

1. Main Database: contains information about which state or jurisdiction uses which tool as well as detailed information about the tools and many other resources.
2. Glossary: an explanation of the terms used.
3. Tools: detailed information about the tools used as well as many other resources.
4. Tools Jurisdictions: a great overview of which tool is implemented, organized per tool and jurisdiction
5. State Tools: tools used organized at the state level.
6. Popular tools with Population: includes population statistics for each county that uses a certain tool.

Complementary to this dataset, there is also a pdf document that contains results from interviews with eighteen jurisdictions.

Source: http://pretrialrisk.com/

6. INSPIRATION

To successfully tackle this challenge, your team will need to combine business acumen with sound technical skills. On which area you mainly focus is completely up to your team. Here, we have listed a couple of starting points to solve the challenge. **Note that these are just guidelines and you can solve the challenge in any way you please.**

Focus on one stakeholder

There are many stakeholders related to this challenge:

US Department of Justice

The US DOJ has to ensure a fair and transparant process for every accused. How does the information presented to you impact the transparancy and fairness of criminal trials?

Non-Profit Organizations

Many NPOs fight for human rights including the right of prisoners and accused persons. Can you build a report that shows these organizations if they should oppose the algorithms?

Our society as a whole

Our society benefits from a just prison system. Offenders of crimes should be incarcenated and innocent people should not have to spend time in jail. What is the societal cost or benefit of using these algorithms?

The academic world

Fairness in Machine Learning is a hot topic in ML research. Will your team manage to apply the methods developed by researchers to assess the fairness of the algorithms?

The individual

People get screened by these algorithms on a daily basis. Can you assess the impact of this on the individual that gets screened and build a case against or in favor of these algorithms?

Developers of the algorithms

Various companies have developed algorithms that can be used in the justice system often with great intentions. Can you help them by evaluating their algorithms and improving them?

Your team can consider any of these (and other!) stakeholders and build a solid case for them. Think about what is important for these stakeholders. What do they have to gain with these algorithms and what are the negative impacts on them? Once you identified these points, start looking for answers in the data we provided.

Don't forget to clearly communicate your stakeholder to the judges!

Useful frameworks: MECE framework (7.1), SWOT analysis (7.2)

Focus on one or two crucial aspects of the challenge

Instead of analyzing all aspects of the problem, it often makes more sense to dive deep into one or two crucial aspects of the problem. What follows is a **non-exhaustive** list of aspects to the problem coupled with some of the ways you can tackle that aspect of the problem.

1. **Are the algorithms unfair?**

This is probably the most important aspect of the challenge that every team should address. The scientific community has spent a lot of time developing methods to assess unfairness. Generally speaking, there are three ways you can assess the unfairness of an algorithm: (1) explore the input variables used in the model, do you think these variables should be used?; (2) explore the output of the model, is the algorithm fair for every race and gender?; and (3) visualize the decisions made by the algorithm.

More concretely:

> Are there any variables used that can amplify existing human biases?
>
> Is the outcome decision of the algorithm fair for all genders and races?
>
> Does the algorithm produce more false positives for one specific race or gender?
>
> ...

Definitions of fairness might vary between teams so make sure to properly communicate what you believe fairness is. Keep in mind the context of the challenge. In most justice systems, everyone is considered innocent until proven guilty. How does this affect your definition of fairness? You can also use more advanced techniques like regression analysis to determine whether or not these algorithms are fair.

Useful resources:

> The Role of Protected Attributes in AI Fairness
>
> Definitions of fairness + solutions to unfairness
>
> Technical definitions of unfairness
>
> Machine Learning Model Fairness in Practice

Useful frameworks: fairness modeling tools (7.3), fairness modeling (workshop)

How do these algorithms help [stakeholder]?

You might come to the conclusion that the algorithms are not biased. In that case, it makes sense to continue with an exploration of how these algorithms help various stakeholders.

Useful frameworks: stakeholders (6), MECE-framework (7.1), SWOT analysis (7.2)

2. **How do these algorithms negatively impact [stakeholder]?**

You might conclude that the algorithms are biased. In that case, it makes sense to continue with an exploration of how these algorithms negatively impact various stakeholders.

Useful frameworks: stakeholders (6), MECE-framework (7.1), SWOT analysis (7.2)

3. **How do these algorithms make decisions?**

Even though these algorithms usually communicate what kind of variables they take into account, they are still not transparent in the way they work. The COMPAS algorithm by example does not release data about the questionnaires nor does it say what weights it gives to every topic questioned in the questionnaire. It can thus be considered a black-box model for most of its users.

By using the variables in the dataset you can nonetheless gain insight into how the algorithm works. Does it mainly focus on criminal history? Or how much importance does it attach to the race of the offender? Does it even take the crime itself into account? Various techniques exist to answer these questions.

Useful resources:

Interpretable Machine Learning

Global Surrogate Models

Introduction to scikit-learn

4. **How can we fix the algorithms?**

You might conclude that the algorithms are biased. In that case, you want to fix the issue of unfairness. How can you do this? Generally speaking, there are three distinct ways that the unfairness can be solved: (1) in pre-processing, (2) during processing, and (3) post-processing. There exist many different techniques so you might want to check out the tools we listed in section 7.3.

Useful resources:

Wikipedia page

Fixing unfairness - Comparative study

End-to-end project

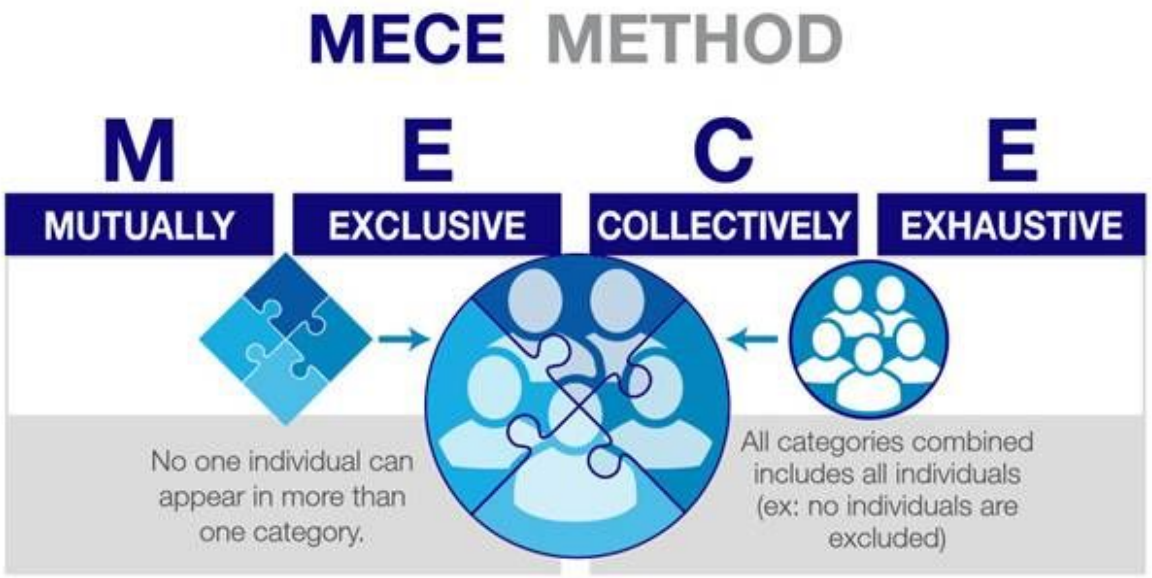
Useful frameworks: fairness modeling tools (7.3)

5. **Any other question you think is important to answer**

7. RELEVANT TOOLS

7.1. MECE-FRAMEWORK

MECE is a principle used by management consulting firms to describe a way of organizing information. The MECE principle suggests that to understand and fix any large problem, you need to understand your options by sorting them into categories that are: **Mutually Exclusive**– Items can only fit into one category at a time and **Collectively Exhaustive** – All items can fit into one of the categories



More information: https://www.caseinterview.com/mece

7.2. SWOT ANALYSIS

SWOT stands for Strengths, Weaknesses, Opportunities, and Threats, and so a SWOT Analysis is a technique for assessing these four aspects. What are the strengths and weaknesses of the current solution (in our case 'using the algorithms') and what are certain threats and opportunities?

Strengths	Weaknesses
What do you do well? What unique resources can you draw on? What do others see as your strength?	What could you improve? Where do you have fewer resources than others? What are others likely to see as a weakness?
Opportunities	**Threats**
What opportunities are open? What trends do you take advantage of? How can you turn your strengths into opportunities?	What threats could harm you? What is your competition doing? What threats do your weaknesses expose you to?

More information: https://www.mindtools.com/pages/article/newTMC_05.htm

7.3. FAIRNESS MODELING TOOLS

Companies have spent a fair amount of resources in creating tools to create fair machine learning models.

Aequitas is an open-source bias and fairness audit toolkit that was released in 2018. It is designed to enable developers to seamlessly test models for a series of bias and fairness metrics concerning multiple population sub-groups.

A Bias and Fairness Audit Toolkit

As part of **Microsoft Fair Learn**, this is a general-purpose methodology for approaching fairness. Using binary classification, the method applies constraints to reduce fair classification to a sequence of cost-sensitive classification problems. Whose solutions yield a randomized classifier with the lowest (empirical) error subject to the desired constraints.

A Reductions approach to fair classification

The **What-if Tool from Google** is an open-source TensorBoard web application that lets users analyze an ML model without writing code. It visualizes counterfactuals so that users can compare a datapoint to the most similar point where the model predicts a different result. Also, users can explore the effects of different classification thresholds, taking into account constraints such as different numerical fairness criteria. There are several demos available – showing how the different functions work on pre-trained models.

The What-if-Tool

IBM 360 degree toolkit contains a comprehensive set of fairness metrics for datasets and machine learning models, explanations for these metrics, and algorithms to mitigate bias in datasets at the pre-processing and model training stages.

The 360-degree toolkit

8. EVALUATION CRITERIA

There are six different prizes that can be won:

1. Best overall team (€1 000)
2. Best data insight (€200)
3. Best technical solution (€200)
4. Best pitch (€200)
5. Best strategy (€200)
6. Best visualization (€200)

The solutions that you present will be judged on many different dimensions. The exact details about how you will be judged will not be shared with the participants. Nonetheless, we have some guidelines that can help you win one of these prizes!

Best data insight	Make sure you present the insights you gained during the challenge in a clear and understandable way. Try to dive deep into the data and find insights that are impactful and free of bias.
Best technical solution	Your solution should be a technical solution that uses the most appropriate techniques to successfully tackle this challenge. You should not be afraid of using more advanced methods as long as they are the right tool for this problem.
Best pitch	Focus on delivering a pitch that can convince stakeholders of your solution. Clearly communicate in a structured and comprehensive way. When one of the judges asks questions, answer them in an insightful way.
Best strategy	Develop a solution that is feasible, valuable and tackles the problem or opportunity you identified in a complete way. Don't forget to clearly communicate what your solution tries to achieve and why it is valuable.
Best visualization	Convince the jury by creating stunning visualizations that contribute to the case you are trying to build. Whether or not you make it interactive is up to you, but it needs to be impactful.
Overall	Find a balance between all the points above and convince the jury that your solution is the solution your stakeholder(s) need(s).

Important note: just like your team the jury consists of technical and non-technical people, make sure you communicate in a way that both of them understand. The jury only has a basic understanding of the topic of the challenge.

9. Submission

In order to be eligible for the prizes your team needs to follow the submission guidelines presented here. **Any violation of these submission guidelines will result in disqualification from winning any of the prizes mentioned above.**

Every team needs to submit a visual representation of what they will present to the jury. In case you only use a PowerPoint presentation it is sufficient to submit the PowerPoint. In case you use any other tool during the presentation, you will have to include screenshots of the tool as well. Most teams will use code. In case your team used any code, this **code also needs to be submitted**. If your team wins a prize, Emergent will check the code for cheating.

The solution you submit needs to be identical to the solution presented. Any deviation will result in disqualification to win any of the prizes. A non-complete list of possible violations: fixing typos, changing the font, changing the colour, adding a slide, removing a slide, changing content of a slide, …

Your pitch can be **7 minutes** long. **Your team is required to share their screen to present the solution.** After your pitch, the jury will have some time to ask questions. Please make sure that your pitch does not exceed the allotted time limit. There will be an Emergent member present that will keep track of your time and cut you off if you go over the time limit.

Submission link:

Submit your files using this link!

Anti-cheating measures

During the creation of the challenge we have implemented anti-cheating measures. During the presentation, an Emergent team member that has been trained to identify cheating will be present. Other anti-cheating measures have been implemented as well but they will not be communicated to the participants.

What is considered cheating

As a general guideline, teams should avoid copying solutions that they found online. Every statistic or figure that can be calculated based on the data has to be created by the team and not taken from the internet. You are allowed to use additional data, resources and figures as long as you clearly state the source of them. In case of doubt, please reach out to an Emergent team member. After four months of working on this challenge, we have gotten a very good idea about what can and cannot be found online. **Cheating will not be tolerated and results in immediate disqualification from the current and all future editions of the D4GC or other competitions hosted by Emergent.**