

Demo

- An early demonstration of the platform is available at: <https://www.youtube.com/watch?v=K6jLhB0MLqk>

Approach

With our industry-focused approach to data science and machine learning, we make it easy for non-technical users—not just developers but also professionals—to get value from their data.

The idea behind HESSIC

As president of PoliMi Data Scientists (the student-led data science association at Politecnico di Milano), I had the opportunity to organize and participate in events of many companies involved in machine learning and data science. One such event that particularly caught my interest was the one we organized at the Milan office of McKinsey & Company.

I was surprised to discover the great effort that data scientists put into making basic tasks such as data cleaning, implementing workflow, and other pre-analysis steps before they can start analyzing their data. The process of cleaning the data alone can take up to 80% of a data scientist's time. This made me think about whether there was a more efficient way to approach data extraction, transformation, and analysis, and a solution that could process the tasks in a fast enough way that it could be ready to be used as a starting point for any project so that data scientists can focus on more complex analysis techniques.

While working in academia, I designed a method for tracking bot-created disinformation websites by sourcing and analyzing tweets from Twitter. The first model took two weeks of development to complete and could not scale well because of the exponential time complexity with respect to the dataset size. I then started looking for better ways to reduce complexity. After several other weeks of researching and implementing, I found a technique called Locality-sensitive hashing. Whereas my previous method took hours to process the data, this new method only took a few seconds. After spending over a month in trial and error, I realized I was facing the same problems as the data scientists at McKinsey: I was wasting too much time implementing basic data pipelines and finding the right algorithm. I asked myself if there was a more efficient way to overcome this bottleneck, but I could not find anything close to my needs. This is where the idea of HESSIC first came to me: a platform to increase the level of data scientists and analysts, not by making them better individually, but by lowering the difficulty of doing complex analysis with data. I thought one could streamline the data analysis process by working on an intuitive and integrated platform for data acquisition and statistical model evaluation. The idea is to make quick work of complex data analysis with a set of basic but powerful tools, so I began developing the first prototype.

After almost two years of development, the next steps will be the most challenging. I will be working to create a scalable architecture that will enable the system to handle increased computing power. Market validation and customer acquisition are both critical from a business standpoint. In the coming months, I am planning to deliver value to non-profit organizations by providing them with free access to my platform.