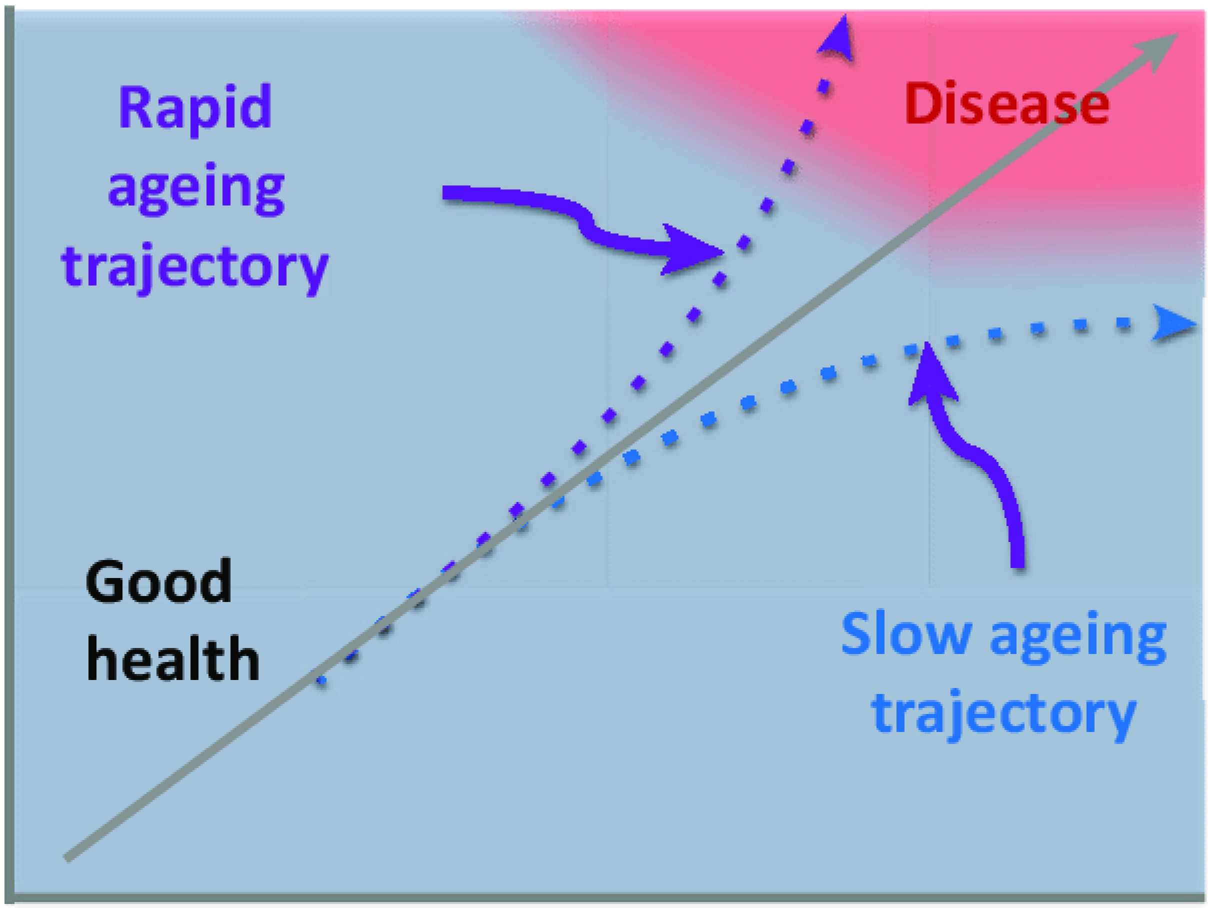
IBM HACK CHALLENGE(2022)

Problem Statement :

Age-related diseases are **killing 150,000 people per day**. Humanity is a health tech organization, which is now able to monitor people's rates of aging, but the only way for that information to have an impact is if the people can know what actions they should take to slow their aging down. This is complex because these impactful actions will not only be different for every person, but also for every moment in that person's life, and for every combination of actions the person takes.

Biological Age: The basic idea behind biological aging is that aging occurs as you gradually accumulate damage to various cells and tissues in the body. Biological age may vary depending on your lifestyle (diet, exercise, sleep, attitude, stress, etc.). Depending on your genetics and your life habits, your biological age will be higher or lower than your chronological one. People with a younger biological age compared to their chronological age are at a lower risk of suffering age-related diseases and mortality.

So in this challenge, Humanity and the Omdena team compressed high throughput markers such as activity and other lifestyle action data from the user (e.g. diet, weight, socio-economic status) to develop weighted algorithms predictive of the biological age outcome.



Importance of Machine learning & AI:

DEFINATION OF MACHINE LEARNING:

Machine learning is a tool that allows systems the ability to learn and improve automatically based upon experience. Machine learning does not need specific programming to carry out an activity. **Machine learning is the development of computer programs** that can access data, and through a series of algorithms use the data to learn for itself what action should be taken based on that data.

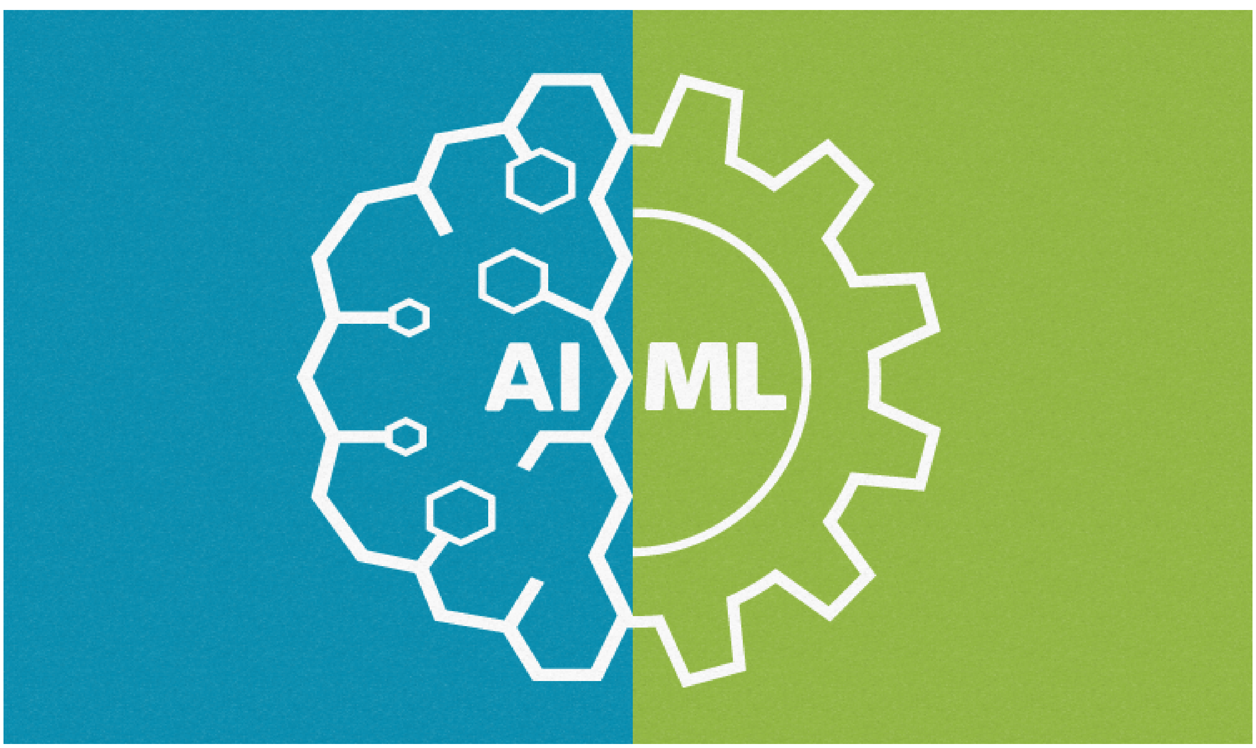
The primary object of machine learning is to allow the system to learn automatically without human intervention. This allows the system to adjust and take action accordingly. The learning process begins with the system observing reference data and experiences based on that data. The system then begins to understand and learn what actions to take when specific patterns within a data set present themselves.

DEFINATION OF AI:

In today's world, technology is growing very fast, and we are getting in touch with different new technologies day by day.

Here, one of the booming technologies of computer science is Artificial Intelligence which is ready to create a new revolution in the world by making intelligent machines.The Artificial Intelligence is now all around us. It is currently working with a variety of subfields, ranging from general to specific, such as self-driving cars, playing chess, proving theorems, playing music, Painting, etc.

Artificial Intelligence is composed of two words **Artificial** and **Intelligence**, where Artificial defines *"man-made,"* and intelligence defines *"thinking power"*, hence AI means *"a man-made thinking power."*



Benifits of AI & ML:

1. Automation

Automation is one of the most commonly cited benefits of AI technology, and it has had significant impacts on the communications, transportation, consumer products, and service industries. Automation not just leads to higher production rates and increased productivity in these sectors but also allows more efficient use of raw materials, improved product quality, reduced lead times, and superior safety. Automation can also help to free resources that can be used for more important things.

2. Smart Decision Making

Artificial Intelligence has always been used for making smarter business decisions. AI technology can coordinate data delivery, analyze trends, develop data consistency, provide forecasts, and quantify uncertainties to make the best decisions for the company. As long as AI is not programmed to imitate human emotions, it will remain unbiased on the matter at hand and will help to make the right decision to support business efficiency.

Artificial Intelligence (AI) has been around for quite some time now. From quick suggestions on search engines and auto-focus in smartphones to robot greeters at shopping centers and vehicle cruise control, AI is increasingly becoming a part of our day-to-day lives. By integrating **AI solutions** into every aspect of the business, organizations can optimize operations, gain a competitive edge and ultimately accelerate growth. The scope for innovation and development in AI is enormous and it will continue changing the world in diverse ways in the future.

Below are the **10 most remarkable benefits of Artificial Intelligence** that are helping to reshape the world that we know of today.

3. Enhanced Customer Experience

AI-powered solutions can help businesses to respond to customer queries and grievances quickly and address the situations efficiently. The use of **chatbots** that couple conversational AI with Natural Language Processing technology can generate highly personalized messages for customers, which helps to find the best solution for their needs. AI tools can also help to reduce the strain from the customer service staff, which will lead to better productivity.

4. Medical Advances

The use of Artificial Intelligence solutions in the **healthcare sector** is becoming increasingly popular these days. Remote patient monitoring technology, for instance, allows healthcare providers to perform clinical diagnoses and suggest treatments quickly without requiring the patient to visit the hospital in-person. AI can also be beneficial in monitoring the progression of contagious diseases and even predict their future effects and outcomes.

5. Research and Data Analysis

AI and **Machine Learning** technology can be used to analyze data much more efficiently.

It can help to create predictive models and algorithms to process data and understand the potential outcomes of different trends and scenarios. Moreover, the advanced computing capabilities of AI can also speed up the processing and analysis of data for research and development, which could have taken too long for humans to review and understand.

6. Solving Complex Problems

The developments in AI technologies from basic Machine Learning to advanced Deep Learning models have made it capable to solve complex issues. From fraud detection and personalized customer interactions to weather forecasting and medical diagnosis, AI is helping businesses across industries to find the right solutions to address their challenges more adequately. Greater efficiency in solving complex problems means increased productivity and reduced expenses.

7. Business Continuity

Business forecasting using AI technology not only helps companies make critical decisions but also prepares them for any emergency to ensure business continuity. As risk management heavily relies on data management and analysis today, AI-powered tools can help organizations to respond to the crisis proactively. AI and Machine Learning can also create scenarios to help businesses plan for a speedy disaster recovery strategy.

8. Managing Repetitive Tasks

Performing recurring business tasks is not just time-consuming but it can also be monotonous and reduce the productivity of the employees over time. AI-powered

Robotic Process Automation tools can automate interactions between different business systems and make the tiresome work easy for the company. It can imitate the actions of humans within the digital systems in the HR, IT, marketing, or sales departments to execute any business process quickly without needing any manual effort.

9. Minimizing Errors

Another great benefit of automating regular business tasks using AI tools is that it helps to reduce the chances of manual errors. As Robotic Process Automation tools take care of the data entry and processing jobs, it can make the digital systems more efficient and less likely to run into or create any problems due to data processing mistakes. This can be especially beneficial for businesses that cannot afford to make even the slightest of errors.

10. Increased Business Efficiency

Artificial Intelligence can help to ensure 24-hour service availability and will deliver the same performance and consistency throughout the day. Taking care of repetitive tasks will not make AI tools get tired or bored either. This can help to improve the efficiency of the business and reduce the stress on the employees, who can be re-assigned to perform more complex business tasks that require manual intervention.

PROJECTDETAILS:

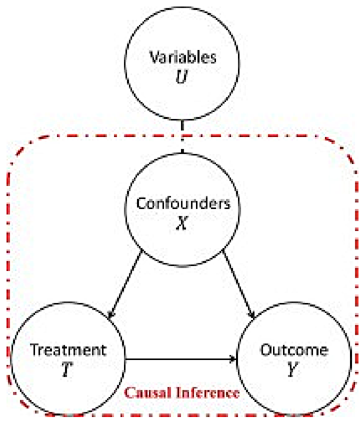
Machine learning algorithms use standard statistical analysis, typified by regression, estimation, and hypothesis testing techniques to assess parameters of a distribution from samples drawn of that distribution. With the help of such parameters, one can infer associations among variables, estimate beliefs or probabilities of past and future events, as well as update those probabilities in light of new evidence or new measurements.

Correlation and causality can seem deceptively similar. While causation and correlation can exist at the same time, correlation does not imply causation. Causation explicitly applies to cases where action A causes outcome B. On the other hand, **correlation is simply a relationship.** Action A relates to Action B — but one event doesn't necessarily cause the other event to happen. So, correlations can lead to wrong assumptions.

In correlations, the notation is P(x|y) i.e. the probability of x given y: for example, the probability of a disease given the person consumes alcohol. However, in causal calculus, a very small but important change is made. **Instead of P(x|y) it's P(x|do(y))** i.e. the probability of x **given that y is done:** for example, the probability of a lower biological age given that I start doing a high-intensity activity. **The 'do' is very important: it represents the intervention, the actual doing of something that will cause the effect.**

Causal inference is a powerful modeling tool for explanatory analysis, which might enable current machine learning to make explainable predictions.

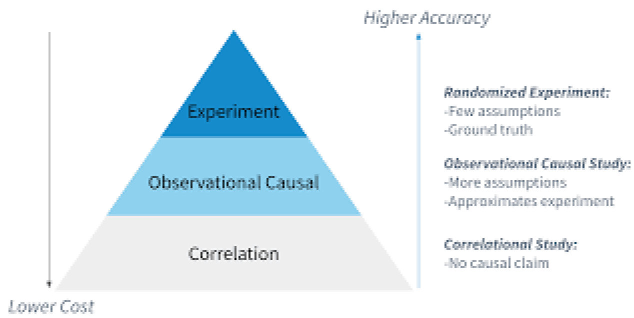
Conceptual Framework (Structural Causal Graphs):



Directed acyclic graphs are used to represent causal relationships. A DAG displays assumptions about the relationship between variables. The assumptions we make take the form of lines (or edges) going from one node to another. Directed paths are also chains because each is causal on the next.

Here, both X and Treatment variables have an impact on biological age. People with lower chronological age or weight might not have a higher biological age. Consequently, their chances of having a disease are lower. Also, people who exercise more, do not consume alcohol, have healthy sleeping patterns may have lower biological age. This is based on the assumption that having healthier lifestyle habits will lead to having a longer life. We are trying to capture the relationship between the treatment i.e. lifestyle actions and outcome variables i.e. biological age.

Simulating randomized controlled trials:

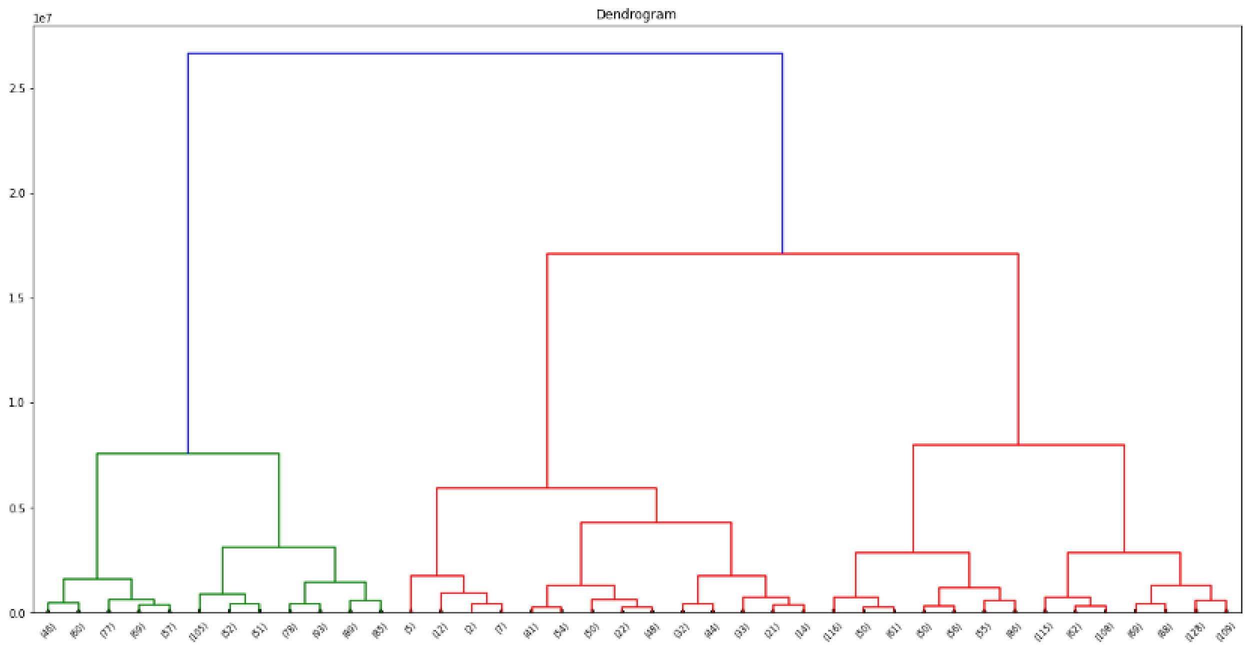


The data that we have used is of 5 people over a period of 6 months recorded from devices like fitbit. Also, there is manually entered data by the user like sleep efficiency, fatigue, stress, revitalization score on a score of 10.

Randomized experiments are the gold standard for causal inference because the treatment assignment is random and physically manipulated: one group gets the treatment, one does not. The assumptions here are straightforward, securable by design, and can be conveniently defended. When there is no control over treatment assignment, like with observational data, we attempt to model it. Modeling here is equivalent to saying "we assume that after adjusting for age, gender, weight, maximum heart rate, alcohol consumption, socioeconomic factors, ethnicity, runners and non-runners are so similar to each other as if they were randomly assigned to running."

Controlled experiments are simple, we can act upon a variable directly and see how our other variables change in our causal diagram. In a medical trial, this would be taking groups of people 1 and 2, 1 group 1 taking the placebo, and group 2 taking the actual medicine to the sickness and observing the results. Naturally, in medical trials we want these people to come from the same distribution.

In order to simulate an RCT experiment environment, the first step will be to cluster similar users based on constant factors that do not change instantly. Hierarchical clustering allows users to move from one cluster to another as the clusters become more mature.

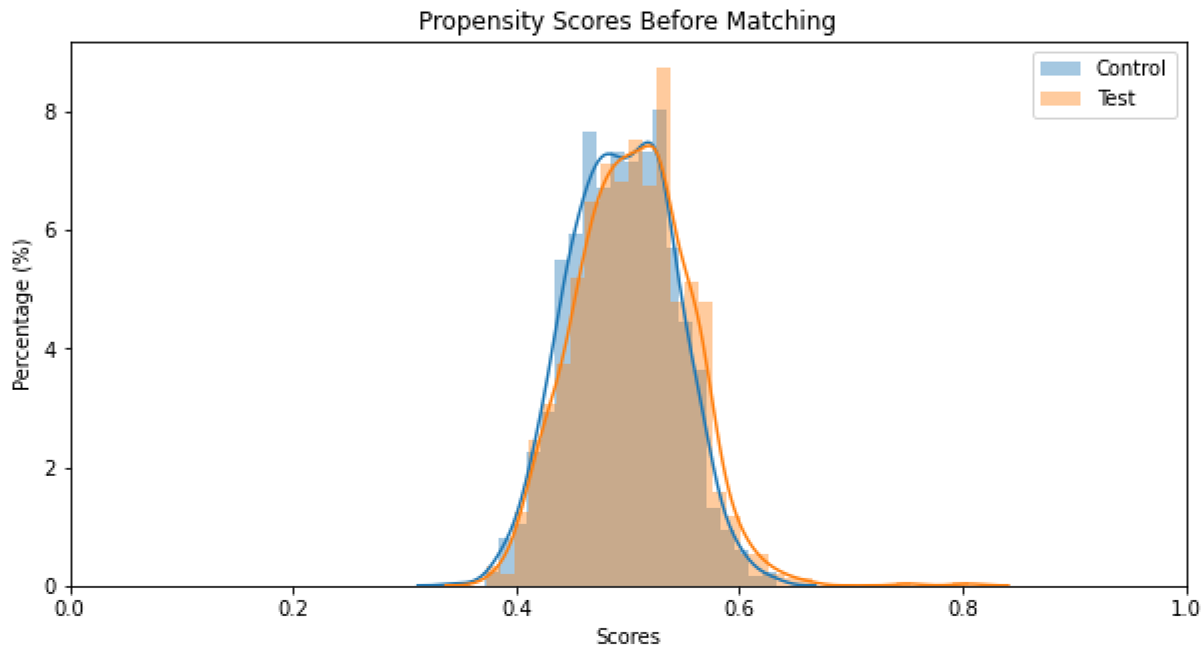


In randomized experiments, treatment is assigned by the flip of a coin, but in observational studies treatment (eg., a person exercising) may be determined by many factors (e.g., likes exercising). If those factors affect the risk of developing the outcome (e.g., lowering of biological age), then the effects of those factors become entangled with the effect of treatment.

So, the next step to mimic controlled experimentation and minimize the limitations of observational data would be to minimize selection bias. For example, in a particular cluster, a certain user tends to exercise more to stay healthy while another user tends to meditate more and sleep for 7–8 hours every day. For this purpose, we have used the propensity score matching technique.

PyMatch is a Python library that features matching techniques for observational studies and is a Python implementation of R's Matching package. PyMatch supports propensity score matching for both discrete and continuous variables, which we used during our project.

In this step, we fitted an initial propensity score model and obtained the following result. We observed that both dummy groups have very similar propensity score distribution. In our case, this is expected since our BA is randomly simulated, and is not affected by any of the activity/ health variables. Ideally (for an accurate dataset), there should be a more visible separation between the 2 groups, thus indicating that the X-variables affect the y-variable, and hence it is worth further implementing matching.



Matching attempts to reduce the treatment assignment bias, and mimic randomization, by creating a sample of units that received the treatment that is comparable on all observed covariates to a sample of units that did not receive the treatment. In this case, matching tries to estimate effects on decreased BA group had they received treatments that are different from the decreased BA group to increased BA group.

Now, we use the function model.match() to start implementing matching. This is done with replacement, meaning a single majority record (in our case, y=1) can be matched to multiple minority records (y=0). Matcher assigns a unique record_id to each record in the test and control groups so this can be addressed after matching. At the end of this section, the scores are printed out and we can observe that each pair of matches have scores within 0.0001 of each other.

Now, we assess the matches by plotting the histograms and ECDF plots of each X-variables. We can observe that for most variables, matching has made the corresponding distribution more similar across the 2 dummy groups. This indicates that the matching algorithm worked as intended.

Machine Learning for Recommendation:

With the matched dataset (Matched_df), we can now train an ML model and infer causality. We used a logistic regression model and performed hyperparameter tuning on its regularization parameter C.

Conclusion

We have built a system that takes in the user actions that are being monitored on one side (activity rates, sleep, meditation, diet, etc.) and uses the ongoing increases or decreases in the user's Rate of Aging measure (to calculate the user's Biological Age) to decide which actions were most effective and in what combinations and when.

The system then also matched across users with similar attributes to use the insights and weightings set for one user to affect the weightings given to actions and the combination of actions to another user. The causal inference has helped us identify the introduction of certain new actions/interventions to a person's daily actions since they have a large effect on the person's rate of aging. Machine learning implementation has allowed the personalization and combinatorial nature of real-life to be modeled.