ABOUT INNOTOMY

Innotomy is the science of exploring existing or new innovations that seek formal approach to derive meaning.

SERVICES

Consulting

Training

Teaching

TECHNOLOGY DOMAINS

Artificial Intelligence

Data Analytics

Data Science

Machine Learning

Deep Learning

1. **TGS Salt Identification Challenge** - Segment salt deposits beneath the Earth's surface.
   Several areas of Earth with large accumulations of oil and gas also have huge deposits of salt below the surface. But unfortunately, knowing where large salt deposits are precisely is very difficult. Professional seismic imaging still requires expert human interpretation of salt bodies. This leads to very subjective, highly variable renderings. More alarmingly, it leads to potentially dangerous situations for oil and gas company drillers. To create accurate seismic images and 3D renderings, TGS (the world's leading geo-science data company) is hoping to build an algorithm that automatically and accurately identifies if a subsurface target is salt or not.
2. **Airbus Ship Detection Challenge** - Find ships on satellite images as quickly as possible.
   Airbus is excited to build a model that detects all ships in satellite images as quickly as possible. Can you find them even in imagery with clouds or haze? Shipping traffic is growing fast. More ships increase the chances of infractions at sea like environmentally devastating ship accidents, piracy, illegal fishing, drug trafficking, and illegal cargo movement. This has compelled many organizations, from environmental protection agencies to insurance companies and national government authorities, to have a closer watch over the open seas. Airbus offers comprehensive maritime monitoring services by building a meaningful solution for wide coverage, fine details, intensive monitoring, premium reactivity and interpretation response. Combining its proprietary-data with highly-trained analysts, they help to support the maritime industry to increase knowledge, anticipate threats, trigger alerts, and improve efficiency at sea.
3. **RSNA Pneumonia Detection Challenge** - Build an algorithm that automatically detects potential pneumonia cases
   The purpose of this challenge is to build an algorithm to detect a visual signal for pneumonia in medical images. Specifically, your algorithm needs to automatically locate lung opacities on chest radiographs. Pneumonia accounts for over 15% of all deaths of children under 5 years old internationally. In 2015, 920,000 children under the age of 5 died from the disease. While common, accurately diagnosing pneumonia is a tall order. It requires review of a chest radiograph (CXR) by highly trained specialists and confirmation through clinical history, vital signs and laboratory exams. Pneumonia usually manifests as an area or areas of increased opacity on CXR. However, the diagnosis of pneumonia on CXR is complicated because of a number of other conditions in the lungs such as fluid overload (pulmonary edema), bleeding, volume loss (atelectasis or collapse), lung cancer, or post-radiation or surgical changes. Outside of the lungs, fluid in the pleural space (pleural effusion) also appears as increased opacity on CXR. When available, comparison of CXRs of the patient taken at different time points and correlation with clinical symptoms and history are helpful in making the diagnosis. CXRs are the most commonly performed diagnostic imaging study.

Deep Learning Use-cases

4. **Digit Recognizer** – Computer vision fundamentals with the famous MNIST data.
   MNIST ("Modified National Institute of Standards and Technology") is the de facto "hello world" dataset of computer vision. Since its release in 1999, this classic dataset of handwritten images has served as the basis for benchmarking classification algorithms. As new machine learning techniques emerge, MNIST remains a reliable resource for researchers and learners alike. In this competition, your goal is to correctly identify digits from a dataset of tens of thousands of handwritten images. We've curated a set of tutorial-style kernels which cover everything from regression to neural networks. We encourage you to experiment with different algorithms to learn first-hand what works well and how techniques compare.
5. **Titanic: Machine Learning from Disaster** - Predict survival on the Titanic.
   The sinking of the RMS Titanic is one of the most infamous shipwrecks in history. On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships. One of the reasons that the shipwreck led to such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others, such as women, children, and the upper-class. This challenge requires one to complete the analysis of which people were likely to survive the tragedy.
6. **Two Sigma: Using News to Predict Stock Movements** - Use news analytics to predict stock price performance.
   Use the content of news analytics to predict stock price performance? The ubiquity of data today enables investors at any scale to make better investment decisions. The challenge is ingesting and interpreting the data to determine which data is useful, finding the signal in this sea of information. As a scientifically driven investment manager, Two Sigma has been applying technology and data science to financial forecasts for over 17 years. Their pioneering advances in big data, AI, and machine learning have pushed the investment industry forward.
7. **House Prices: Advanced Regression Techniques** - Predict sales prices and practice feature engineering, RFs, and gradient boosting.
   Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But this playground competition's dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence. With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, this competition challenges you to predict the final price of each home.

ABOUT INNOTOMY

Innotomy is the science of exploring existing or new innovations that seek formal approach to derive meaning.

SERVICES

Consulting
Training
Teaching

TECHNOLOGY DOMAINS

Artificial Intelligence
Data Analytics
Data Science
Machine Learning
Deep Learning

innotomy
DATA SCIENCE CONSULTING

Deep Learning Use-cases

8. **New York City Taxi Fare Prediction** - Can you predict a rider's taxi fare?
   In this playground competition, hosted in partnership with Google Cloud and Coursera, you are tasked with predicting the fare amount (inclusive of tolls) for a taxi ride in New York City given the pickup and dropoff locations. While you can get a basic estimate based on just the distance between the two points, this will result in an RMSE of $5-$8, depending on the model used (see the starter code for an example of this approach in Kernels). Your challenge is to do better than this using Machine Learning techniques!
9. **Forest Cover Type** - Use cartographic variables to classify forest categories
   In this competition you are asked to predict the forest cover type (the predominant kind of tree cover) from cartographic variables. The actual forest cover type for a given 30 x 30 meter cell was determined from US Forest Service (USFS) Region 2 Resource Information System data. Independent variables were then derived from data obtained from the US Geological Survey and USFS. The data is in raw form and contains binary columns of data for qualitative independent variables such as wilderness areas and soil type. This study area includes four wilderness areas located in the Roosevelt National Forest of northern Colorado. These areas represent forests with minimal human-caused disturbances, so that existing forest cover types are more a result of ecological processes rather than forest management practices.
10. **Movie Review Sentiment Analysis** - Classify the sentiment of sentences from the Rotten Tomatoes dataset
    "There's a thin line between likably old-fashioned and fuddy-duddy, and The Count of Monte Cristo ... never quite settles on either side."

    The Rotten Tomatoes movie review dataset is a corpus of movie reviews used for sentiment analysis, originally collected by Pang and Lee. In their work on sentiment treebanks, Socher et al used Amazon's Mechanical Turk to create fine-grained labels for all parsed phrases in the corpus. This competition presents a chance to benchmark your sentiment-analysis ideas on the Rotten Tomatoes dataset. You are asked to label phrases on a scale of five values: negative, somewhat negative, neutral, somewhat positive, as positive. Obstacles like sentence negation, sarcasm, terseness, language ambiguity, and many others make this task very challenging.
11. **TrackML Particle Tracking Challenge** - High Energy Physics particle tracking in CERN detectors
    To explore what our universe is made of, scientists at CERN are colliding protons, essentially recreating mini big bangs, and meticulously observing these collisions with intricate silicon detectors. While orchestrating the collisions and observations is already a massive scientific accomplishment, analyzing the enormous amounts of data produced from the experiments is becoming an overwhelming challenge. Event rates have already reached hundreds of millions of collisions per second, meaning physicists must sift through tens of petabytes of data per year. And, as the resolution of detectors improve, ever better software is needed for real-time pre-processing and filtering of the most promising events.

Deep Learning Use-cases

12. **Home Credit Default Risk** - Can you predict how capable each applicant is of repaying a loan?
    Many people struggle to get loans due to insufficient or non-existent credit histories. And, unfortunately, this population is often taken advantage of by untrustworthy lenders. Home Credit strives to broaden financial inclusion for the unbanked population by providing a positive and safe borrowing experience. In order to make sure this underserved population has a positive loan experience, Home Credit makes use of a variety of alternative data--including telco and transactional information--to predict their clients' repayment abilities. While Home Credit is currently using various statistical and machine learning methods to make these predictions, they're challenging you to help them unlock the full potential of their data. Doing so will ensure that clients capable of repayment are not rejected and that loans are given with a principal, maturity, and repayment calendar that will empower their clients to be successful.
13. **2018 Data Science Bowl** - Find the nuclei in divergent images to advance medical discovery
    Imagine speeding up research for almost every disease, from lung cancer and heart disease to rare disorders. The 2018 Data Science Bowl offers our most ambitious mission yet: create an algorithm to automate nucleus detection. We've all seen people suffer from diseases like cancer, heart disease, chronic obstructive pulmonary disease, Alzheimer's, and diabetes. Many have seen their loved ones pass away. Think how many lives would be transformed if cures came faster. By automating nucleus detection, you could help unlock cures faster—from rare disorders to the common cold.
14. **Plant Seedlings Classification** - Determine the species of a seedling from an image
    Can you differentiate a weed from a crop seedling? The ability to do so effectively can mean better crop yields and better stewardship of the environment. The Aarhus University Signal Processing group, in collaboration with University of Southern Denmark, has recently released a dataset containing images of approximately 960 unique plants belonging to 12 species at several growth stages.
15. **Dog Breed Identification** - Determine the breed of a dog in an image
    Who's a good dog? Who likes ear scratches? Well, it seems those fancy deep neural networks don't have all the answers. However, maybe they can answer that ubiquitous question we all ask when meeting a four-legged stranger: what kind of good pup is that?

    In this playground competition, you are provided a strictly canine subset of ImageNet in order to practice fine-grained image categorization. How well you can tell your Norfolk Terriers from your Norwich Terriers? With 120 breeds of dogs and a limited number training images per class, you might find the problem more, err, ruff than you anticipated.

Deep Learning Use-cases

16. **Mercari Price Suggestion Challenge** - Can you automatically suggest product prices to online sellers?
    Product pricing gets even harder at scale, considering just how many products are sold online. Clothing has strong seasonal pricing trends and is heavily influenced by brand names, while electronics have fluctuating prices based on product specs. Mercari, Japan's biggest community-powered shopping app, knows this problem deeply. They'd like to offer pricing suggestions to sellers, but this is tough because their sellers are enabled to put just about anything, or any bundle of things, on Mercari's marketplace. In this competition, Mercari's challenging you to build an algorithm that automatically suggests the right product prices. You'll be provided user-inputted text descriptions of their products, including details like product category name, brand name, and item condition.
17. **Statoil/C-CORE Iceberg Classifier Challenge** - Ship or iceberg, can you decide from space?
    Drifting icebergs present threats to navigation and activities in areas such as offshore of the East Coast of Canada. Currently, many institutions and companies use aerial reconnaissance and shore-based support to monitor environmental conditions and assess risks from icebergs. However, in remote areas with particularly harsh weather, these methods are not feasible, and the only viable monitoring option is via satellite. Statoil, an international energy company operating worldwide, has worked closely with companies like C-CORE. C-CORE have been using satellite data for over 30 years and have built a computer vision based surveillance system. To keep operations safe and efficient, Statoil is interested in getting a fresh new perspective on how to use machine learning to more accurately detect and discriminate against threatening icebergs as early as possible. In this competition, you're challenged to build an algorithm that automatically identifies if a remotely sensed target is a ship or iceberg. Improvements made will help drive the costs down for maintaining safe working conditions.
18. **Planet: Understanding the Amazon from Space** - Use satellite data to track the human footprint in the Amazon rainforest
    Every minute, the world loses an area of forest the size of 48 football fields. And deforestation in the Amazon Basin accounts for the largest share, contributing to reduced biodiversity, habitat loss, climate change, and other devastating effects. But better data about the location of deforestation and human encroachment on forests can help governments and local stakeholders respond more quickly and effectively. Planet, designer and builder of the world's largest constellation of Earth-imaging satellites, will soon be collecting daily imagery of the entire land surface of the earth at 3-5 meter resolution. While considerable research has been devoted to tracking changes in forests, it typically depends on coarse-resolution imagery from Landsat (30 meter pixels) or MODIS (250 meter pixels). This limits its effectiveness in areas where small-scale deforestation or forest degradation dominate.

ABOUT INNOTOMY

Innotomy is the science of exploring existing or new innovations that seek formal approach to derive meaning.

SERVICES

Consulting
Training
Teaching

TECHNOLOGY DOMAINS

Artificial Intelligence
Data Analytics
Data Science
Machine Learning
Deep Learning

innotomy
DATA SCIENCE CONSULTING

Deep Learning Use-cases

19. **NOAA Fisheries Steller Sea Lion Population Count** - How many sea lions do you see?
Steller sea lions in the western Aleutian Islands have declined 94 percent in the last 30 years. The endangered western population, found in the North Pacific, are the focus of conservation efforts which require annual population counts. Specially trained scientists at NOAA Fisheries Alaska Fisheries Science Center conduct these surveys using airplanes and unoccupied aircraft systems to collect aerial images. Having accurate population estimates enables us to better understand factors that may be contributing to lack of recovery of Stellers in this area. Currently, it takes biologists up to four months to count sea lions from the thousands of images NOAA Fisheries collects each year. Once individual counts are conducted, the tallies must be reconciled to confirm their reliability. The results of these counts are time-sensitive. In this competition, you are invited to develop algorithms which accurately count the number of sea lions in aerial photographs. Automating the annual population count will free up critical resources allowing NOAA Fisheries to focus on ensuring we hear the sea lion's roar for many years to come. Plus, advancements in computer vision applied to aerial population counts may also greatly benefit other endangered species.
20. **Dstl Satellite Imagery Feature Detection** - Can you train an eye in the sky?
As these large, complex datasets continue to increase exponentially in number, the Defence Science and Technology Laboratory (Dstl) is seeking novel solutions to alleviate the burden on their image analysts. In this competition, Kagglers are challenged to accurately classify features in overhead imagery. Automating feature labeling will not only help Dstl make smart decisions more quickly around the defense and security of the UK, but also bring innovation to computer vision methodologies applied to satellite imagery.
21. **Data Science Bowl 2017** - Can you improve lung cancer detection?
In the United States, lung cancer strikes 225,000 people every year, and accounts for $12 billion in health care costs. Early detection is critical to give patients the best chance at recovery and survival. One year ago, the office of the U.S. Vice President spearheaded a bold new initiative, the Cancer Moonshot, to make a decade's worth of progress in cancer prevention, diagnosis, and treatment in just 5 years. In 2017, the Data Science Bowl will be a critical milestone in support of the Cancer Moonshot by convening the data science and medical communities to develop lung cancer detection algorithms. Using a data set of thousands of high-resolution lung scans provided by the National Cancer Institute, participants will develop algorithms that accurately determine when lesions in the lungs are cancerous. This will dramatically reduce the false positive rate that plagues the current detection technology, get patients earlier access to life-saving interventions, and give radiologists more time to spend with their patients. You are called to observe the right patterns, ask the right questions, and in turn, create unprecedented impact around cancer screening care and prevention.

ABOUT INNOTOMY

Innotomy is the science of exploring existing or new innovations that seek formal approach to derive meaning.

SERVICES

Consulting
Training
Teaching

TECHNOLOGY DOMAINS

Artificial Intelligence
Data Analytics
Data Science
Machine Learning
Deep Learning

innotomy
DATA SCIENCE CONSULTING

Deep Learning Use-cases

22. **Dogs vs. Cats** - Distinguish images of dogs from cats
    In 2013, we hosted one of our favorite for-fun competitions: Dogs vs. Cats. Train your neural network on host of cats and dogs images and then predict in an unknown image of cat or dog, which one is which.
23. **Leaf Classification** - Can you see the random forest for the leaves?
    There are estimated to be nearly half a million species of plant in the world. Classification of species has been historically problematic and often results in duplicate identifications. Automating plant recognition might have many applications, including:

    - Species population tracking and preservation
    - Plant-based medicinal research
    - Crop and food supply management

    The objective of this playground competition is to use binary leaf images and extracted features, including shape, margin & texture, to accurately identify 99 species of plants. Leaves, due to their volume, prevalence, and unique characteristics, are an effective means of differentiating plant species. They also provide a fun introduction to applying techniques that involve image-based features.
24. **Integer Sequence Learning** - 1, 2, 3, 4, 5, 7?!
    You read that correctly. That's the start to a real integer sequence, the powers of primes. Want something easier? How about the next number in 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55? If you answered 89, you may enjoy this challenge. Your computer may find it considerably less enjoyable. The On-Line Encyclopedia of Integer Sequences is a 50+ year effort by mathematicians the world over to catalog sequences of integers. If it has a pattern, it's probably in the OEIS, and probably described with amazing detail. This competition challenges you create a machine learning algorithm capable of guessing the next number in an integer sequence. While this sounds like pattern recognition in its most basic form, a quick look at the data will convince you this is anything but basic!
25. **Ultrasound Nerve Segmentation** - Identify nerve structures in ultrasound images of the neck
    Even the bravest patient cringes at the mention of a surgical procedure. Surgery inevitably brings discomfort, and oftentimes involves significant post-surgical pain. Currently, patient pain is frequently managed through the use of narcotics that bring a bevy of unwanted side effects. This competition's sponsor is working to improve pain management through the use of indwelling catheters that block or mitigate pain at the source. Pain management catheters reduce dependence on narcotics and speed up patient recovery. Accurately identifying nerve structures in ultrasound images is a critical step in effectively inserting a patient's pain management catheter. In this competition, Kagglers are challenged to build a model that can identify nerve structures in a dataset of ultrasound images of the neck. Doing so would improve catheter placement and contribute to a more pain free future.

Deep Learning Use-cases

26. **BNP Paribas Cardif Claims Management** - Can you accelerate BNP Paribas Cardif's claims management process?
    As a global specialist in personal insurance, BNP Paribas Cardif serves 90 million clients in 36 countries across Europe, Asia and Latin America. In a world shaped by the emergence of new uses and lifestyles, everything is going faster and faster. When facing unexpected events, customers expect their insurer to support them as soon as possible. However, claims management may require different levels of check before a claim can be approved and a payment can be made. With the new practices and behaviors generated by the digital economy, this process needs adaptation thanks to data science to meet the new needs and expectations of customers. In this challenge, BNP Paribas Cardif is providing an anonymized database with two categories of claims: claims for which approval could be accelerated leading to faster payments claims for which additional information is required before approval
27. **Prudential Life Insurance Assessment** - Can you make buying life insurance easier?
    Prudential, one of the largest issuers of life insurance in the USA, is hiring passionate data scientists to join a newly-formed Data Science group solving complex challenges and identifying opportunities. The results have been impressive so far but we want more. In a one-click shopping world with on-demand everything, the life insurance application process is antiquated. Customers provide extensive information to identify risk classification and eligibility, including scheduling medical exams, a process that takes an average of 30 days. The result? People are turned off. That's why only 40% of U.S. households own individual life insurance. Prudential wants to make it quicker and less labor intensive for new and existing customers to get a quote while maintaining privacy boundaries.
28. **Cervical Cancer Screening** - Help prevent cervical cancer by identifying at-risk populations
    Cervical cancer can be prevented through early administration of the HPV vaccine and regular pap smear screenings, which indicate the presence of precancerous cells. It is also sometimes curable by the removal of the early-stage cancerous tissue that is identified through pap smears. Screening and early treatment can lead to potential cures in about 95% of women at risk for cervical cancer. There are many patient advocacy programs on the importance of pap smears in cervical cancer prevention. However, these widespread programs may not be reaching or effectively speaking to the most vulnerable populations. If one could better identify these women, education campaigns could target them with content that speaks directly to their unique risk factors. Identifying predictors of not receiving pap smears will provide important information to stakeholders in cervical cancer prevention who run awareness programs. Genentech is asking you to join their mission to help prevent cervical cancer. Given a dataset of de-identified health records, your challenge is to predict which women will not be screened for cervical cancer on the recommended schedule.

Deep Learning Use-cases

29. **Rossmann Store Sales** - Forecast sales using store, promotion, and competitor data
    Rossmann operates over 3,000 drug stores in 7 European countries. Currently, Rossmann store managers are tasked with predicting their daily sales for up to six weeks in advance. Store sales are influenced by many factors, including promotions, competition, school and state holidays, seasonality, and locality. With thousands of individual managers predicting sales based on their unique circumstances, the accuracy of results can be quite varied. In their first Kaggle competition, Rossmann is challenging you to predict 6 weeks of daily sales for 1,115 stores located across Germany. Reliable sales forecasts enable store managers to create effective staff schedules that increase productivity and motivation. By helping Rossmann create a robust prediction model, you will help store managers stay focused on what's most important to them: their customers and their teams!
30. **Diabetic Retinopathy Detection** - Identify signs of diabetic retinopathy in eye images
    Currently, detecting DR is a time-consuming and manual process that requires a trained clinician to examine and evaluate digital color fundus photographs of the retina. By the time human readers submit their reviews, often a day or two later, the delayed results lead to lost follow up, miscommunication, and delayed treatment. Clinicians can identify DR by the presence of lesions associated with the vascular abnormalities caused by the disease. While this approach is effective, its resource demands are high. The expertise and equipment required are often lacking in areas where the rate of diabetes in local populations is high and DR detection is most needed. As the number of individuals with diabetes continues to grow, the infrastructure needed to prevent blindness due to DR will become even more insufficient. The need for a comprehensive and automated method of DR screening has long been recognized, and previous efforts have made good progress using image classification, pattern recognition, and machine learning. With color fundus photography as input, the goal of this competition is to push an automated detection system to the limit of what is possible – ideally resulting in models with realistic clinical potential.
31. **West Nile Virus Prediction** - Predict West Nile virus in mosquitoes across the city of Chicago
    West Nile virus is most commonly spread to humans through infected mosquitoes. Around 20% of people who become infected with the virus develop symptoms ranging from a persistent fever, to serious neurological illnesses that can result in death. Given weather, location, testing, and spraying data, this competition asks you to predict when and where different species of mosquitoes will test positive for West Nile virus. A more accurate method of predicting outbreaks of West Nile virus in mosquitoes will help the City of Chicago and CPHD more efficiently and effectively allocate resources towards preventing transmission of this potentially deadly virus.

Deep Learning Use-cases

32. **Traffic sign Recognition** - Train and identify German traffic signs images
    The "German Traffic Sign Recognition Benchmark" is a multi-category classification competition held at IJCNN 2011. Automatic recognition of traffic signs is required in advanced driver assistance systems and constitutes a challenging real-world computer vision and pattern recognition problem. A comprehensive, lifelike dataset of more than 50,000 traffic sign images has been collected. It reflects the strong variations in visual appearance of signs due to distance, illumination, weather conditions, partial occlusions, and rotations.
33. **Credit card fraud detection** - Predict fraudulent transaction
    Annual global fraud losses reached $21.8 billion in 2015, according to Nilson Report. Probably you feel very lucky if you are a fraud. About every 12 cents per $100 were stolen in the US during the same year. Train a neural network in unsupervised (or semi-supervised) manner for anomaly detection in credit card transaction data.
34. **How much did it rain?** - Predict probabilistic distribution of hourly rain given polarimetric radar measurements
    For agriculture, it is extremely important to know how much it rained on a particular field. However, rainfall is variable in space and time and it is impossible to have rain gauges everywhere. Therefore, remote sensing instruments such as radar are used to provide wide spatial coverage. Rainfall estimates drawn from remotely sensed observations will never exactly match the measurements that are carried out using rain gauges, due to the inherent characteristics of both sensors. Currently, radar observations are "corrected" using nearby gauges and a single estimate of rainfall is provided to users who need to know how much it rained. This competition will explore how to address this problem in a probabilistic manner. Knowing the full probabilistic spread of rainfall amounts can be very useful to drive hydrological and agronomic models -- much more than a single estimate of rainfall. You are given polarimetric radar values and derived quantities at a location over the period of one hour. You will need to produce a probabilistic distribution of the hourly rain gauge total.
35. **LANL earthquake prediction** - Can you predict upcoming laboratory earthquakes?
    Forecasting earthquakes is one of the most important problems in Earth science because of their devastating consequences. Current scientific studies related to earthquake forecasting focus on three key points: when the event will occur, where it will occur, and how large it will be. You will address when the earthquake will take place. Specifically, you'll predict the time remaining before laboratory earthquakes occur from real-time seismic data.
36. **Higgs Boson machine learning challenge** - Use the ATLAS experiment to identify the Higgs boson
    The goal of the Higgs Boson Machine Learning Challenge is to explore the potential of advanced machine learning methods to improve the discovery significance of the experiment. Using simulated data with features characterizing events detected by ATLAS, your task is to classify events into "tau tau decay of a Higgs boson" versus "background."

Deep Learning Use-cases

37. **2016 Data Science Bowl** - Transforming How We Diagnose Heart Disease
    We all have a heart. Although we often take it for granted, it's our heart that gives us the moments in life to imagine, create, and discover. Yet cardiovascular disease threatens to take away these moments. Each day, 1,500 people in the U.S. alone are diagnosed with heart failure—but together, we can help. We can use data science to transform how we diagnose heart disease. By putting data science to work in the cardiology field, we can empower doctors to help more people live longer lives and spend more time with those that they love. You are required to create an algorithm to automatically measure end-systolic and end-diastolic volumes in cardiac MRIs. You will examine MRI images from more than 1,000 patients. This data set was compiled by the National Institutes of Health and Children's National Medical Center and is an order of magnitude larger than any cardiac MRI data set released previously.
38. **Malaria detection** - Detection of malaria from a blood sample
    Malaria is a life-threatening disease caused by parasites that are transmitted to people through the bites of infected female Anopheles mosquitoes. It is preventable and curable. Malaria parasites can be identified by examining under the microscope a drop of the patient's blood, spread out as a "blood smear" on a microscope slide. Prior to examination, the specimen is stained to give the parasites a distinctive appearance. This technique remains the gold standard for laboratory confirmation of malaria. However, it depends on the quality of the reagents, of the microscope, and on the experience of the laboratorian.
39. **Genomic visualization** - Genomic Visualization via Dimensionality Reduction
    Embed and visualize your 23andMe genotype data using samples from the 1000 Genomes Project as a reference population. Ancestry-informative snps (single nucleotide polymorphisms) are locations in the genome that have significant variance across global populations. There are several scientific publications which have shared these locations and this kernel uses 55 AISNPs from Kidd et al. Dimensionality redcution techniques include PCA, t-SNE, and UMAP.
40. **Skin lesion analyzer** – Classify skin lesions
    Skin cancer is a major public health problem with over 123,000 newly diagnosed cases worldwide in each year. Skin cancer is the most common form of cancer, globally accounting for at least 40% of cases. Skin cancer is diagnosed visually, beginning with an initial clinical screening and followed potentially by dermoscopic analysis, a biopsy and histopathological examination. Automated classification of skin lesions using images is a challenging task owing to the fine-grained variability in the appearance of skin lesions. The HAM10000 dataset consists of 10015 dermatoscopic images which are released as a training set for academic machine learning purposes and are publicly available through the ISIC archive. This benchmark dataset can be used for machine learning and for comparisons with human experts.

Deep Learning Use-cases

41. **Bone age estimation** - Estimate chronological age of children between 10-20 years with hand X-rays
    It is important to verify the chronological age of a child between 10 and 20 years of age. A birth certificate and/or parental testimony is simply not acceptable. TW2 and TW3 are the currently recommended techniques of estimating bone age from hand X-rays. However, there are limited number of experts who are certified to conduct these tests and is simply not scalable. This project explores the use of deep learning to train a neural network to estimate bone age given hand X-rays of children between 10 and 20 years of age and compare that estimation with TW3 results. The deep learning approach is expected to hold a lot of promise.

ABOUT INNOTOMY

Innotomy is the science of exploring existing or new innovations that seek formal approach to derive meaning.

SERVICES

Consulting
Training
Teaching

TECHNOLOGY DOMAINS

Artificial Intelligence
Data Analytics
Data Science
Machine Learning
Deep Learning

innotomy
DATA SCIENCE CONSULTING