Project Data & AI – Project 1

Client name:	Inholland University of Applied Sciences, Robotics department
Project name:	Facial recognition and pain detection in Stroke patients with partial paralysis using deep learning

About the client:

The robotics department of Inholland University of Applied sciences focuses on a variety of application ranging from agricultural picking machines to healthcare robots. Within this department a group of researchers and students are working on ISHA, a robot designed to aid physical rehabilitation in stroke patients at home.

About the project:

The robot ISHA is designed to instruct stroke patients to perform arm exercises while seated. It is known that these exercises may lead to discomfort. In a setting with a physiotherapist, this can readily be detected. In a home setting with a care giver but without a physiotherapist, discomfort must be detect based on the patient's facial expression.

Emotion and pain detection using facial images is not a new problem. However, our target population differs considerably from the general population. This is expressed by a more generally depressed state, mostly elderly individuals, and paralysis of one side of the body and face. It is unclear if models trained on healthy individuals are sensitive enough to detect emotions and pain in this population.

The main research questions are as follows:

- *Is it possible to perform facial recognition on images of stroke patients with partial paralysis?*
- *What are the requirements and performance indicators of a model for detection of emotions and pain in stroke patients with partial facial paralysis?*

About the dataset:

There is ample data available of emotions and pain in healthy individuals and certain patient groups. The data related to faces with partial paralysis in stroke is very limited. Initially, we will provide data of semi-paralyzed faces that can be compared to healthy faces for the first research question.

As specific images of pain and discomfort are not readily available for our target population, we would like the project team to simulate the dataset based on datasets of healthy individuals and images of stroke patients.

Available data:

- Stroke patients vs. healthy (https://www.kaggle.com/datasets/danish003/face-images-of-acute-stroke-and-non-acute-stroke)
- Pain recognition from facial expression database(https://www.kaggle.com/datasets/coder98/emotionpain)

- Emotion and identity detection from face images (https://www.kaggle.com/c/facial-keypoints-detector/data)
- Pain e-motion database: Pain E-Motion Faces Database 279 images from 68 individuals
- Multimodal pain database (https://vap.aau.dk/mintpain-database/)
- BioVid heat pain database (https://www.nit.ovgu.de/BioVid.html)
- Affective facial expression database (https://www.affectiva.com/facial-expression-dataset/)

Contact details:

Alistair Vardy – alistair.vardy@inholland.nl
Margo van Kemenade - margo.vankemenade@inholland.nl

Project Data & AI – Project 2

Client name:	RIPE NCC (www.ripe.net)
Project name:	Predicting Internet latency

About the client:

The RIPE NCC is an independent, not-for-profit organisation that works to support the Internet in our service region (Europe, Middle East and parts of Asia) and beyond. The RIPE NCC supports the infrastructure of the Internet through technical coordination in our service region. Our most prominent activity is to act as the Regional Internet Registry (RIR) providing global Internet resources and related services (IPv4, IPv6 and AS Number resources) to members in our service region.

More info at: www.ripe.net / https://www.ripe.net/about-us/what-we-do/

About the project:

With RIPE Atlas we collect the latencies between measurement points on the Internet. In our RIPE Atlas anchors mesh, we do this between a few hundred servers world-wide. So for each pair of measurement points (we call them RIPE Atlas anchors) we have a time series of latencies. We would like to have an accurate prediction of the latency and latency variation in the upcoming 15 minutes between a pair of measurement points, based on the recent past (for instance the last 7 days' worth of data, or the last 4 weeks).
Our question: Is AI able to make better predictions on this data than traditional statistical methods, like Holt Winters forecasting.

Ideally, the deliverable for this project is a code repo with python code that can answer these questions

About the dataset:

An example of the data is available here:
This is what the data looks like for a few sources in Taiwan to a destination in the USA:
https://atlas.ripe.net/api/v2/measurements/50894291/results/?start=1735732800&stop=1735992000&probe_ids=6700,6995,6907,6818,7356,7230&format=json

(each source has a different probe ID (prb_id), for different destinations you need a different destination ID).
API description:
RIPE Atlas APIs and other documentation available at: https://atlas.ripe.net/docs/apis/rest-api-manual/

Contact details:

Emile Aben (emile.aben@ripe.net)

Project Data & AI – Project 3

Client name:	TNO
Project name:	Classifying Arrhythmia in Fighter Pilots

About the client:

TNO's mission is to generate innovative solutions with demonstrable impact to achieve a safe, healthy, sustainable, and digital society and boost the earning power of the Netherlands. We aim to achieve this by performing two core tasks (roles). The first is to support the Dutch government in carrying out statutory government tasks in the public interest. TNO's second core task is to strengthen the earning power of the Dutch economy and increase employment.

About the project:

This project focuses on identifying unexplained physiological events (UPEs) experienced by F-35 pilots during high-stress, high-G maneuvers. Pilots sometimes report feeling unwell during these maneuvers, raising the critical question: is this due to a genuine physiological anomaly, or is it a false alarm? To address this, we aim to develop analytical modules that process ECG signals from sensors worn by the pilots.

The primary goal is to design a robust arrhythmia classifier capable of detecting irregular heartbeats, such as premature or delayed heart contractions. However, the project faces a significant challenge: ensuring the classifier remains effective across different domains and sensor types. Given that the specific ECG sensor and measurement methods might vary in the future, the algorithm must generalize well to unseen signal sources. Furthermore, there exists no data of pilots becoming with many arrhythmic heart beats during flight (luckily), and therefore any classifier has to be trained on data from controlled experiments, but should work during flight. This is a second reason why generalization is important.

Initial experiments suggest that traditional models such as Random Forests perform well on the dataset they are trained on but fail on new datasets with different characteristics. Conversely, Support Vector Classifiers (SVCs) show better cross-domain robustness. You will work on improving domain transfer robustness through innovative methodological choices.

About the dataset:

To facilitate development, students will use a labeled ECG dataset available from PhysioNet (https://www.physionet.org/content/mitdb/1.0.0/), featuring arrhythmias recorded from individuals at rest. This supervised training phase provides a foundation for the classifier. The next step involves evaluating the model on a second dataset, where participants engage in different activities such as running or cycling—mimicking the high-variability conditions of fighter pilots. This second dataset may lack arrhythmia labels, requiring students to devise novel evaluation strategies, such as manual labeling or unsupervised approaches. We will select a fitting dataset together, there are multiple options on PhysioNet.

Contact details:

Maarten.schadd@tno.nl

Project Data & AI – Project 4

Client name:	KPN BV
Project name:	AI Enhanced Quality Assurance

About the client:

KPN is the Dutch leading telecommunication and ICT service provider for consumer, business and wholesale markets. At KPN we are all passionate about offering secure, reliable and future-proof networks and services, enabling people to be connected anytime, anywhere, whilst at the same time creating a more prosperous and cleaner world. We have been doing this based on a strong vision, for almost 150 years. KPN is listed on the Euronext stock exchange. More on us: https://overons.kpn/

About the project:

When bringing fiber connections to households, KPN uses 3rd-parties to install a Fiber Termination Unit (FTU) within the customer premises, very often in the proximity of the energy meter in the meter cupboard (meterkast). The positioning of the FTU is regulated by well-defined quality standards which the 3rd-party should meet. Unfortunately, such standards are not met in a relevant number of installations, generating fiber provisioning delays, customer unsatisfaction and additional costs.

The main deliverable of this project is a computer vision model which is able to detect an (in)correctly installed FTU – i.e. according to the required quality standards - at the customer premises by processing a picture of such installation.

A good performing model could then enable the development of a production grade application (out-of-scope) which will be used by the 3rd party field engineers to certify correct FTU installation to KPN.

To further improve the business value of such project, we encourage the participation of multiple student groups, possibly building additional models (object detection, region proposal network, bounding box regression, spatial reasoning, etc.) able to suggest upfront the ideal location on where to install the FTU, by using a (synthetic) picture of a candidate spot of the customer premises.

About the dataset:

The training and validation datasets will jointly contain at least 1000 datapoints, i.e. pictures of FTU installations at customer premises, to be labeled as (non-)compliant with the help of KPN Quality Assurance experts and conform the current installation quality criteria.

An exploration dataset is already available and you will use it to select the useful images, i.e. those where it is clear whether at least one of the quality standard is (un)met, in order to precisely define how to take the pictures to be included in the training set and create a complete distribution of good/bad installations, based on all the quality criteria.

A field trip is planned to experience the (non-)compliant FTU installations yourself, to ask the experts anything and even to help taking the pictures and gathering customer consent.

Contact details:

I'm happy to be your contact on this project. My name is Eric Oostendorp.
Email: eric.oostendorp@kpn.com Mobile: +31 6 22 814 815
Looking forward to meet and work with you!

Project Data and AI – Project 5

Client name	DDSS and RR Mechatronics
Project name:	Multiple Sclerosis blood samples analysis

About the client:

The Data-Driven Smart Society research group of the Research & Innovation Centre for Technology and Computer Science (RIC-TOI) focuses mainly on data analysis, particularly practice-oriented research into artificial intelligence (AI), such as neural networks. DDSS bridges the connection between fundamental data science research and innovative practice applications. Maintaining close ties with universities and knowledge institutions, DDSS collaborates through the multidisciplinary lecturers' platform PRIO for practice-oriented ICT research. Their work in practice-oriented research is inherently multidisciplinary: possessing comprehensive knowledge of data, DDSS collaborates with other experts who have specific information requests and understand which data can offer insights. Depending on the issue, DDSS partners with stakeholders in various sectors, including manufacturing, healthcare, government, and education. DDSS also delves into learning analytics, exploring how data can be leveraged to provide students with even more tailored guidance.

About the project:

Multiple Sclerosis (**MS**) is a chronic autoimmune disease that affects the central nervous system and affects over 2.8 million people worldwide. Emerging research highlights the potential role of red blood cells (RBCs) in MS, mainly through iron deposits or oxidative stress.

1. Clustering and Dimensionality Reduction

Unsupervised Learning: Clustering algorithms like K-means, DBSCAN, or Gaussian Mixture Models (GMM) can be used to group MS and healthy samples based on their underlying patterns. Visualize high-dimensional features using techniques like t-SNE, UMAP, or PCA to find meaningful separations between MS and healthy groups.

Objectives: Explore natural groupings in the data. Identify potential subtypes of MS based on blood sample features.

2. Anomaly Detection

Approach: Train a model on healthy control samples and use it to detect anomalies in MS patient samples.

Algorithms: Autoencoders, Isolation Forest, One-Class SVM.

Objective: Identify subtle deviations in blood samples that may indicate early signs of MS or variations in the disease.

About the dataset:

The dataset comprises two main folders: one folder contains healthy control RBCs, and the other contains 10 folders of images of MS patients' blood samples.

Contact details:

DDSS Researcher: **Dr. Noor Christoph**: noor.christoph@inholland.nl
RR Mechatronics: **Jan Zoeten**: jan.dezoeten@rrmechatronics.com

Project Data and AI – Project 6

Client name:	DataDriven Smart Society(DDSS), Inholland
Project name:	Sustainable Building: Analyzing Mycellium Growth

About the client:

The Data-Driven Smart Society research group of the Research & Innovation Centre for Technology and Computer Science (RIC-TOI) focuses mainly on data analysis, particularly practice-oriented research into artificial intelligence (AI), such as neural networks. DDSS bridges the connection between fundamental data science research and innovative practice applications. Maintaining close ties with universities and knowledge institutions, DDSS collaborates through the multidisciplinary lecturers' platform PRIO for practice-oriented ICT research. Their work in practice-oriented research is inherently multidisciplinary: possessing comprehensive knowledge of data, DDSS collaborates with other experts who have specific information requests and understand which data can offer insights. Depending on the issue, DDSS partners with stakeholders in various sectors, including manufacturing, healthcare, government, and education. DDSS also delves into learning analytics, exploring how data can be leveraged to provide students with even more tailored guidance.

About the project:

Due to escalating demands, the construction industry faces mounting pressure to integrate non-conventional materials. Traditional materials such as timber, masonry, fiber cement, and metal struggle to keep up with global population growth. They also cause adverse environmental impacts. One solution is to adopt circular construction practices focused on reducing waste, promoting recycling, and using sustainable materials — biodegradable glass, for example, or recycled tires.

Another innovative alternative gaining traction involves using mycelium, the root structure of fungi, as a construction material for bricks and cladding. Mycelium-based materials are biodegradable, consume little energy, and have a low carbon footprint. The client has time-series data (unstructured) of mycelium growth acquired for 10-14 days. The objective is to deploy an unsupervised machine learning algorithm to analyze whether groups of images could be clustered together having similar features across time. If the clustering by the model captures features across the timeline, it will help to identify the development phases of the mycelium.

About the dataset:

The available dataset has multiple folders of mycelium growth images taken across the timestamp. The total number of images is more than 5000.

Contact details:

DDSS Researcher:
Tilman Koster: tilmann.koster@inholland.nl
Twan Defesche: twan.defesche@inholland.nl

Project Data and AI – Project 7

Client name:	Professorship Robotics | Smart Farming, Inholland
Project name:	Qualification and segmentation of chicory pens

About the client:

At the Professorship Robotics | Smart Farming at the University of Applied Sciences Inholland Alkmaar, we facilitate innovation in agricultural robotics by leading cutting-edge applied research in a team with industry, researchers, and students, thereby solving real-life challenges. We focus on robotics in pome fruits (e.g., apples, pears), field vegetables (e.g., broccoli, strawberries, chicory), and – a bit of a Dutch pride - flower bulbs.

About the project:

Chicory, or Brussels chicory, is a leafy vegetable grown biennially. In spring, chicory is sown directly in the soil. The seeds grow into large green endive-like heads, with thick carrot-like brown-white "pens" developing underground. In autumn, these pens are harvested and (without leaves) brought to the chicory farm. There, they are kept in cold storage for up to a year. Periodically, new trays filled with pens are placed in a controlled environment and grown in water over a few weeks ("forcing chicory") until they develop into full white heads.

Placing the pens in the trays is a very labor-intensive process. In this project, we're exploring to what extent this process can be automated. As such, the pens need to be detected and segmented, and the top, bottom, and contact points need to be identified.

About the dataset:

The dataset consists of 1018 high-resolution images, with 78 images annotated using bounding boxes. The annotations include object dimensions such as length and width and labels for each object. The dataset captures objects like roots and stems in various orientations, shapes, and sizes, ensuring diversity. All images are of high quality, captured under controlled conditions for consistency.

Contact details:

Robotics Research Group
Associate Lector: Dr. Jeroen Wildenbeest (jeroen.wildenbeest@inholland.nl)
Researcher: Kristel van Ammers (kristel.vanammers@inholland.nl) & Surya Giri (surya.giri@inholland.nl)

Project Data and AI – Project 8

Client name:	Professorship Robotics | Smart Farming, Inholland
Project name:	Counting roses

About the client:

At the Professorship Robotics | Smart Farming at the University of Applied Sciences Inholland Alkmaar, we facilitate innovation in agricultural robotics by leading cutting-edge applied research in a team with industry, researchers, and students, thereby solving real-life challenges. We focus on robotics in pome fruits (e.g., apples, pears), field vegetables (e.g., broccoli, strawberries, chicory), and – a bit of a Dutch pride - flower bulbs.

About the project:

Aalsmeer Flower Auction (Bloemenveiling Aalsmeer) is a flower auction in Aalsmeer, Netherlands. It is the largest flower auction in the world. The Aalsmeer Flower Auction building has the ninth-largest by-floor area in the world, covering 999,000 square meters (10,750,000 sq ft; 247 acres). Flowers from all over the world are traded here every day; around 43 million flowers are sold daily. Greenport Aalsmeer, among other things, centralizes the innovation at Bloemenveiling Aalsmeer.

One of the flowers being traded is roses. These roses arrive in carton boxes, bundled in sets of 8-12. They are unpacked, checked for diseases, and hung in a sorting machine. Greenport Aalsmeer wishes to automate this process. As such, the roses in the bundles need to be counted and checked for diseases.

About the dataset:

The dataset consists of images extracted from GoPro video footage, captured at 30 frames per second to ensure sufficient detail and information. It contains approximately 5000 raw, unlabeled images. Since the dataset is unprocessed, it requires cleaning, preprocessing, and enhancement before it can be used for analysis or model training.

Contact details:

Robotics Research Group
Associate Lector: Dr. Jeroen Wildenbeest (jeroen.wildenbeest@inholland.nl)
Researcher: Kristel van Ammers (kristel.vanammers@inholland.nl) & Surya Giri (surya.giri@inholland.nl)

Project Data and AI – Project 9

Client name:	Research group RIC TOI – Inholland, World Horti Center, Lentiz group.
Project name:	Detection of Tomato Head thickness Diameter from RGB-D Camera Images Using Computer Vision and Deep Learning Techniques in Greenhouse Conditions

About the client:

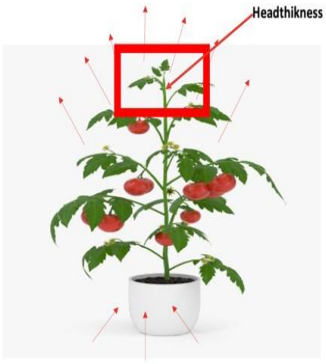
The Research and Innovation Centre Techniek, Ontwerpen & Informatica (RIC-TOI) of Inholland University of Applied Sciences contributes to a sustainable living environment and resilient society. Whether it's the increasing pressure on healthcare, the call for efficient and ecological food production or the challenges surrounding climate change: today's issues require solutions that are smart. Practically applicable, without creating new problems. Solutions, therefore, in which effective technological innovations play a key role.

About the project:

The world's population is predicted to reach 9.1 billion by 2050, and this will require an increase of 70% in food availability, but so far 1 billion people are still suffering from hunger. It was estimated that 30% of the food produced for human consumption globally is lost or wasted somewhere along the food supply chain. Food loss and waste have significant negative food-security, economic and environmental impacts. Tackling food losses and waste in an efficient, sustainable, and integrated way should be considered an urgent thing to feed people and optimize the use of natural and financial resources at the same time.

For growers, head thickness is a crucial indicator of tomato plant health, with an ideal target measurement of approximately 10mm. A thickness below this target suggests weak plant growth, making the plants more vulnerable to pests and diseases. On the other hand, if the thickness exceeds this threshold, it indicates that the plant is diverting excess sugars to its leaves instead of focusing on flower and fruit development. Accurately detecting and managing head thickness is therefore essential for optimizing plant health and yield in greenhouse conditions.

The primary objective of this assignment is to develop a system that accurately detects and measures the head thickness diameter of tomatoes in a greenhouse environment. This system will utilize RGB-D cameras to capture depth information, and leverage Machine Learning, Deep Learning, and Computer Vision techniques to process the data and perform the measurements.



1. **How can RGB-D camera images be utilized to accurately measure the head thickness diameter of tomatoes in greenhouse condi5ons?**
   What preprocessing steps are necessary to enhance the quality of RGB-D data for tomato measurement?
2. **Which computer vision techniques are most effective for segmenting and detecting tomato heads in RGB-D images under varying greenhouse conditions? e.g 3D-Cloud? YOLO v9,10,11**
   How do lighting conditions, occlusions, and plant positioning affect the segmentation accuracy?
3. **What deep learning models perform best for es5ma5ng tomato head thickness diameter from RGB-D images?**
   How do different neural network architectures (e.g., CNNs, U-Net) compare in terms of accuracy and computational efficiency?
4. **How does the inclusion of depth informa5on from RGB-D cameras improve the accuracy of tomato head thickness measurements compared to using RGB images alone?**
   What role does depth data play in distinguishing tomato head features that are critical for diameter estimation?
5. **What are the challenges of implementing this detection system in real-time greenhouse environments, and how can they be addressed?**
   How robust is the system against common greenhouse variables such as fluctuating light conditions, temperature changes, and varying plant growth stages?
6. **What are the poten5al benefits and limita5ons of using deep learning and computer vision for agricultural applica5ons, specifically for measuring tomato head thickness in greenhouse conditions?**
   What impact does this approach have on labor efficiency, accuracy of crop monitoring, and overall greenhouse management?

About the dataset:

Images data set. Images collected from RGBD camera, which is a depth camera that provides both depth (D) and color (RGB) data as the output in real-time.

Contact details:

Ir. Amora Amir
Amora.amir@inholland.nl

Project Data and AI – Project 10

Client name:	Incision
Project name:	Comparing and optimizing materials needed for surgical procedure

About the client:

Incision is a company that was founded 10 years ago by Theo Wiggers, a surgeon. Incision started as an e-learning company for surgical professionals. Nowadays Incision supports operating room (OR) workflows using an app named Assist. This app contains all relevant information needed for a surgical team to be well prepared for a procedure in the OR: It contains information for setting of the OR, the devices, the preparation of the patient, the materials needed and the steps of the procedure, including the preferences of each surgeon in that specific hospital.

About the project:

Ideally surgeons use the same materials (instruments and disposables) for their procedures, but there is a variety in what they use: This lack of standardization may lead to delays (team does not have the right materials), mistakes (team gives the wrong material) and costs (using materials that are more expensive than needed). During this project you identify the variety in materials used and come up with a dashboard for the OR manager to show the room for improvement and can give an advice how to standardize their practice

About the dataset:

When we look at 1 hospital we do have about 600 procedures, each procedure containing between 5 or 10 surgeons performing this procedure, so 3000-6000 lists of materials. For each procedure a surgeon would need between 20 and 40 materials. Of these materials 90% will be similar for each surgeon, but the interesting part is in the varying 10%

Contact details:

Marian Scheer
Scheer@incision.care
06-44557296

Project Data and AI – Project 11

Client name:	Moodies Undies
Project name:	Life cycle value x Moodies Undies

About the client:

Moodies undies is a Dutch, Amsterdam-based scale-up. Period products create a lot of disposable waste, and these products are not always comfortable. There are also all kinds of chemicals in these products that are not at all good for your body's health. In addition, for young girls it is not at all nice and easy to use. There is also still a taboo on menstruation and many women and girls know little about their own body. There is also still enough menstrual poverty in the Netherlands. This is why moodies is committed to offering a comfortable, stylish and sustainable alternative. In this way we try to reconnect women with their cycle and make them aware that menstruation is part of your overall health. Our dream is to substantially reduce menstrual waste. Moodies pants can make a significant contribution to this. We will all have to work together towards a cleaner planet for our children in which we make more sustainable choices to leave the planet healthier for our children. And we are really happy that we can offer many women and girls a great solution. For girls and ladies with a heavier menstruation, for girls who have their first period and for ladies who experience mild incontinence.

About the project:

The main problem we want to have solved is the minimal understanding of our customers. We know who they are and what they like but we want to study their behavior. We want to know that when they order from us, when they will order again. But also, where did they come from, direct, search, social media. In the end the conclusion of the project will tell us which fields we need to focus on and when we can retarget the people that recently bought something in our store.

About the dataset:

Our store is built in Shopify. The back end tracks data from customers and tells you if they have bought before. We can export data that shows first the customers vs returning customers, separate documents of returning customers and first time customers that didn't return (so far;)) and we can export all the orders. When we export we can deliver a CSV document.

Contact details:

If you want to reach us there are a few ways
Claire - owner: claire@moodiesundies.com
Bente - marketing project manager: bente@moodiesundies.com
General: info@moodiesundies.com (this is mostly if one of the above doesn't respond)

Project Data and AI – Project 12

Client name:	MEDxAI
Project name:	Diagnostic Support for cardiologists

About the client:

MEDxAI is a AI startup in healthcare. MEDxAI specializes in data science challenges in healthcare. The amount of healthcare data is growing to staggering amounts and the potential of exploiting this data to improve healthcare is widely recognized. However, it remains difficult to harness this potential for the benefit of patients and the amount of data driven diagnostic support systems a doctor and patient can use in the clinical practice remains limited to this day. Our mission is to bring data science into the doctor's office and to assist the doctor in giving the best possible care to the patient.

About the project:

For the development of a clinical decision support system we extract model features from electronic health records from a hospital. The data contained in the electronic health records span multiple data modalities, ranging from structured tabular data like laboratory measurements to unstructured timeseries data such as ECG data and free text data. Each modality requires a bespoke extraction approach.

Specifically for processing radiology reports we want to estimate the probability that a radiology report is associated with certain medical conditions.

This is challenging because the Dutch clinical language is under-resourced in terms of Natural Language Processing (NLP) tooling and hence this assignment may lead to a significant contribution to this field.

You are tasked with:

- preparing the data
- model selection
- model training
- deployment via a REST server
- *and writing clean/reproducible code*

About the dataset:

We will have a semi-public research dataset containing annotated translated English radiology reports, with, in the order of hundreds of thousands of documents.

Contact details:

Tjebbe Tauber
CEO MEDxAI
Tjebbe@medxai.com
0627190280

Project Data and AI – Project 13

Client name:	Hearteye
Project name:	Diagnostic model on ECG data from a novel device

About the client:

HeartEye was founded by Dutch cardiologists who realized that the majority of deaths due to cardiovascular diseases happen outside the hospital. The main goal is helping patients to reduce hospital visits. The solution we found is developed in a pocket size ECG monitor which can be connected to the HeartEye app. This recorder simplifies the measurement procedure to a level that is workable for everyone, so patients can acquire their own ECG. The second step is to automate the interpretation using AI so that not only doctors but patients as well can monitor their heart rhythm anywhere and anytime.

About the project:

There is a large ECG dataset available made with traditional ECG recorders (that differs slightly from our recordings). This is the main dataset. Second, a novel ECG recorder has been developed and a prototype is available. With this recorder you create the second (small) dataset. During this project you will use the large public data set and pre-train a model that can diagnose ECGs. Second, you will finetune that model on the small dataset, made with the Hearteye ECG recorder.

You are tasked with:

1) Pre-train a multi-channel time-series classifier on annotated semi-public ECG data
   a) Use Neurokit to derive relevant intervals for ECG's (R-R interval, Q-T time, etc)
2) Gather data with Hearteye Recorder (e.g. from yourself, recorder is supplied by Hearteye)
   a) Create synthetic arrhythmia data
3) Finetune your model on Hearteye data (alternative: check the validity of the model trained on public data)
4) Test the validity of you arrhythmia detection model.
   - *writing clean/reproducible code*
   - *(optional) deployment via a REST server*

About the dataset:

The public ECGs come from physionet. You can register at https://physionet.org/register/ . There are ~800.000 ECGs from 160.000 patients, with annotations (explaining what is the diagnosis of the ECG). Look at:

- MIMIC IV ECG
- MIMIC IV ECG annotated
- https://neuropsychology.github.io/NeuroKit/examples/ecg_delineate/ecg_delineate.html
- https://pure.tue.nl/ws/portalfiles/portal/314931961/mbe-20-11-848.pdf

You could make a selection of normal ECGs and ECGs with arrhythmia. You create the second dataset with the hearteye recorder.

Contact details:

Tjebbe Tauber, CEO Hearteye
Tjebbe@hearteye.nl
0627190280

Project Data & AI – Project 14

Client name:	Heemskerkse Golfclub
Project name:	**SmartSwing** – De kracht van AI die onze golfbaan slimmer maakt.

About the client:

Wij zijn de Heemskerkse Golfclub, een actieve en gastvrije golfvereniging met een prachtige 18-holes baan en een uitdagende Par3-baan. Met een levendig ledenbestand, gastspelers en regelmatige wedstrijden en evenementen, zijn wij dagelijks bezig om een optimale ervaring te bieden aan iedereen die van golf houdt.
Om onze banen efficiënt te beheren en onze leden en gasten goed van dienst te zijn, maken wij gebruik van het reserveringssysteem Egolf4U. Dit systeem bevat alle gegevens over ons ledenbestand en wordt gebruikt voor het plannen van starttijden, zowel voor leden als gastspelers, maar ook voor wedstrijden en evenementen.

About the project:

Hoewel Egolf4U ons voorziet van waardevolle gegevens, halen we deze op dit moment handmatig uit het systeem. Een administratrice verzamelt de benodigde informatie en presenteert dit in spreadsheets. Dit proces is tijdsintensief, beperkt de snelheid van onze besluitvorming en maakt het lastig om trends en patronen in de baanbezetting te herkennen.

We willen een volgende stap zetten door onze data beter te benutten en te koppelen aan externe data zoals bijvoorbeeld het KNMI. Historische gegevens over reserveringen bevatten potentieel waardevolle inzichten die ons kunnen helpen om:

* Voorspellingen te doen over de baanbezetting, bijvoorbeeld met behulp van AI-modellen.
* Betere beslissingen te nemen over marketingactiviteiten, zoals het aantrekken van gastspelers of het promoten van rustige periodes.
* Informatie toegankelijker te maken, bijvoorbeeld in de vorm van interactieve dashboards, zodat we snel en eenvoudig overzicht hebben.

About the dataset:

In Excel (.xls/.csv). Hoeveel data is ons niet bekend. De data betreft o.m. speeltijden, speelfrequenties, no-shows, persoonsinfo e.m.

Contact details:

Joost Bollen, Commercieel manager
joost@heemskerksegolfclub.nl
06 20606345

Project Data & AI – Project 15

Client name:	SAIL Amsterdam
Project name:	Sales & revenue prediction for 2030

About the client:

Since 1975, Tall Ships from around the world have come to Amsterdam every five years. SAIL was first organized 50 years ago by the Port of Amsterdam as a celebration of the city's 700th anniversary. Due to its immense success, it was decided to repeat the event every five years. Over the years, SAIL has grown into the largest free-access event in the Netherlands.

Since 1980, the Foundation SAIL Amsterdam has successfully organized all editions every five years. Well, almost all. As you may know, the 2020 edition was unfortunately canceled due to Covid. Fortunately, preparations for 2025 are already in full swing. The organization aims to make SAIL Amsterdam the most impactful event ever; after all, 2025 marks the 50th anniversary of SAIL and the 750th anniversary of Amsterdam.

Want to get an idea of how grand and special SAIL is? Check out the after movie (https://www.youtube.com/watch?v=J9zEvgR6_F0) of SAIL Amsterdam 2015.

About the project:

The sales of hospitality packages are an important source of revenue for the SAIL event. In the past this was done by an external company, but in 2013 the SAIL organization decided to take this task upon itself for the edition of 2015. Prediction of sales of these packages are important for budget predictions, but also difficult, due to the fact that the event only takes place every 5 years and the world and also systems that can be worked with change considerably in the meantime.

After SAIL Amsterdam 2015 sales predictions were made for the edition of 2020. However, the system used in 2015 only recorded actual sales, not the quotation. Also the numbers had to be normalized for various 'incidents' like taking over bookings from another booking agency.

As mentioned before SAIL Amsterdam 2020 could unfortunately not take place due to Covid and although packages were sold before that, the sales numbers were only reliable up until January 2024 to make predictions for the edition of 2025. Next to that, the organization has a feeling that that booking patterns might have changed: large groups for example have other booking characteristics than other segments, smaller groups tend to book later, but probably more than in previous years.

At this moment SAIL is working hard on selling as many packages as possible for the event in August 2025. Next to that the hospitality department is already looking forward to 2030. How can the sales & revenue of SAIL Amsterdam 2030 be predicted? Therefore we would like you to analyze the current and previous sales revenue numbers to make a sales revenue prediction model for SAIL Amsterdam 2030.

About the dataset:

We have sales data from 2015. For the years 2020 and 2025, we have data from the moment of request to booking. This means we have conversion rates and can better estimate how many

requests for quotations are needed at different times. We have approximately 3000 sales opportunities (requests for quotations) and 1500 bookings. Most of these are business-related, so the dataset can be enriched with characteristics of these companies (sector, size, number of branches, etc.)
Contact details:
Jasper Stikkelorum jasper@sail.nl 06-14062224