

List of Projects

1. **Project Title:** Predictive maintenance for industrial equipment using sensor data and machine learning.

Source: [Microsoft Azure Predictive Maintenance](#)

Description:

This project aims to develop a predictive maintenance system for industrial equipment using sensor data and machine learning algorithms. The goal is to reduce downtime and maintenance costs by predicting equipment failures before they occur. The project involves collecting sensor data from industrial equipment, preprocessing the data, and developing a machine learning model to predict equipment failures. The model will be trained on historical data and tested on new data to evaluate its accuracy. The project will also involve optimizing maintenance schedules based on the predictions made by the model. The dataset includes:

- Telemetry time series data – It consists of hourly average of voltage, rotation, pressure, vibration collected from 100 machines for the year 2015.
- Errors / faults History – These are errors encountered by the machine while in operating conditions. These errors do not shut down or cause failure of the machines.
- Maintenance History – If a component of a machine is replaced, that is captured as a record in this table.
- Machine Features – The features of the machine, e.g., engine size, make and model, location.

Project Outcomes:

- Developed a machine learning model to predict equipment failures.
- Analyzed sensor data to identify patterns and anomalies that indicate impending equipment failure.
- Optimized maintenance schedules to minimize downtime and reduce costs.

2. Project Title: Optimizing energy usage in buildings using IoT data and deep learning algorithms.

Source: [Individual Household Electric Power Consumption](#)

Description:

This project aims to develop a machine learning-based system to optimize energy usage in buildings using data from IoT sensors. The goal is to reduce energy consumption and costs while maintaining occupant comfort and health. The project involves preprocessing data from IoT sensors, developing a deep learning model to optimize energy usage, and testing the model on new data to evaluate its effectiveness. The project will also involve developing user interfaces to enable building managers and occupants to monitor and control energy usage in real-time. The dataset includes:

- Measurements of electric power consumption in one household with a one-minute sample rate over a period of almost 4 years.
- Different electrical quantities and sub-metering values are available.
- The dataset contains 2075259 measurements gathered in a house located in Sceaux (7km of Paris, France) between December 2006 and November 2010 (47 months).
- The dataset contains missing values in measurements (nearly 1.25% of the rows). All calendar timestamps are present in the dataset but for some timestamps, the measurement values are missing.

Project Outcomes:

- Developed a deep learning model to optimize energy usage in buildings.
- Analyzed sensor data to identify patterns and trends in energy usage.
- Developed user interfaces to enable building managers and occupants to monitor and control energy usage in real-time.

3. Project Title: Real-time traffic prediction and optimization for smart cities using sensor data and machine learning.

Source: [Small City Traffic Patterns](#)

Description:

You are working with the government to transform your city into a smart city. The vision is to convert it into a digital and intelligent city to improve the efficiency of services for the citizens. One of the problems faced by the government is traffic. You are a data scientist working to manage the traffic of the city better and to provide input on infrastructure planning for the future.

The government wants to implement a robust traffic system for the city by being prepared for traffic peaks. They want to understand the traffic patterns of the four junctions of the city. Traffic patterns on holidays, as well as on various other occasions during the year, differ from normal working days. This is important to take into account for your forecasting. The dataset includes:

- Training data measurements of the number of vehicles crossing four intersections (junctions) with a one-hour sample rate over a period of almost 2 years (from 2017-11-01 to 2017-06-30 to 2017-10-31).
- Test data is from 2017-06-30 to 2017-10-31.
- A sample of the training data is as follows:

Date Time	Junction	Vehicles	ID
2015-11-01 00:00:00	1	15	20151101001
2015-11-01 01:00:00	1	13	20151101011
2015-11-01 02:00:00	1	10	20151101021
2015-11-01 03:00:00	1	7	20151101031
2015-11-01 04:00:00	1	9	20151101041
2015-11-01 05:00:00	1	6	20151101051
2015-11-01 06:00:00	1	9	20151101061
2015-11-01 07:00:00	1	8	20151101071

Project Outcomes:

- Develop a machine learning model to reduce traffic congestion.
- Analyze sensor data to identify patterns and trends in a junction usage.
- Improve transportation efficiency and enhanced urban mobility.

4. **Project Title:** Environmental Conditions Detection

Source: [Environmental Sensor Telemetry Data](#)

Description:

The data was generated from a series of three identical, custom-built, breadboard-based sensor arrays. Each array was connected to a Raspberry Pi devices. Each of the three IoT devices was placed in a physical location with varied environmental conditions The dataset includes:

Each IoT device collected a total of seven different readings from the four sensors on a regular interval. Sensor readings include temperature, humidity, carbon monoxide (CO), liquid petroleum gas (LPG), smoke, light, and motion. The data spans the period from 07/12/2020 00:00:00 UTC – 07/19/2020 23:59:59 UTC. There is a total of 405,184 rows of data.

There are nine columns in the dataset.

column	description	units
ts	timestamp of event	epoch
device	unique device name	string
co	carbon monoxide	ppm (%)
humidity	humidity	percentage
light	light detected?	boolean
lpg	liquid petroleum gas	ppm (%)
motion	motion detected?	boolean
smoke	smoke	ppm (%)
temp	temperature	Fahrenheit

In this project students need to detect environmental conditions according to IoT sensor collected data.

Project Outcomes:

- Load and analyze the dataset, visualize the dataset and data distribution.
- Pre-processing the dataset.
- Build and design different machine learning models and FNN model.
- Use all features to train the models and predict the device / environmental condition.
- Find out using which features / data combination can get the best environment conditions detection model.

5. Project Title: Appliances Energy Prediction

Source: [Appliances Energy Prediction Dataset](#)

Description:

Experimental data used to create regression models of appliances energy use in a low energy building.

The data set is at 10 min for about 4.5 months. The house temperature and humidity conditions were monitored with a ZigBee wireless sensor network. Each wireless node transmitted the temperature and humidity conditions around 3.3 min. Then, the wireless data was averaged for 10 minutes periods. The energy data was logged every 10 minutes with m-bus energy meters. Weather from the nearest airport weather station (Chievres Airport, Belgium) was downloaded from a public data set from Reliable Prognosis (rp5.ru) and merged together with the experimental data sets using the date and time column. Two random variables have been included in the data set for testing the regression models and to filter out non predictive attributes (parameters).

There are 29 columns in the dataset. Some columns are:

Column	Description	Units
Date	time year-month-day	hour:minute:second
Appliances	energy use	Wh
Lights	energy use of light fixtures in the house	Wh
T1-T9	Temperature in different area	Celsius
RH1-RH9	Humidity in different area	%
To	Temperature outside	Celsius
Pressure	Pressure outside	mm Hg
RHout	Humidity	outside
Wind speed	Wind speed outside	m/s
Visibility	Visibility outside	km
Tdewpoint	Dew point	Celsius
rv1	Random variable 1	nondimensional
rv2	Random variable 2	nondimensional

In this project students need to predict the usage of energy according to WSN sensor collected data.

Project Outcomes:

- Load and analyze the dataset, visualize the dataset and data distribution.
- Pre-processing the dataset. Build and design different machine learning models and FNN model, if time permits, try RNN model.
- Find out using which features / data combination can get the best environment conditions detection model.

6. **Project Title:** Induction Motor Fault Detection

Source: [Machinery Fault Dataset](#)

Description:

One of the most common causes of mechanical breakdown is an industrial motor failure. Companies often rely on motors' day-in-and-day-out to keep uptime going at industrial plants. Over time, slight imbalances in the system can develop into vibrations that wear down the engine and eventually lead to failure. Monitoring motor vibrations are a popular and effective method of looking for signs of neglect and even predicting motor failure before it occurs. Using the dataset included in this project proposal, your group will develop a model to predict the likelihood of motor failure.

This database comprises 1951 multivariate time-series acquired by sensors on SpectraQuest's Machinery Fault Simulator (MFS) Alignment-Balance-Vibration (ABVT). The data consists of six different simulated states: normal function, imbalance fault, horizontal and vertical misalignment faults, and inner & outer bearing defects.

The data was acquired using eight variety sensors to monitor vibration using acceleration, rotation, and sound. Data from these eight sensors were collected at varying load weights and fault states.

More details regarding data structure and organization can be found on the dataset [website](#).

Project Outcomes:

- Use the normal and imbalanced data of the induction motor to detect anomalies in motor function.
- Train a model to predict the likelihood of motor failure at a specific point of time in the continuous data.

7. **Project Title:** Driver Drowsiness Detection

Source: [Drowsiness Detection](#)

Description:

A self-driving car or autonomous car is a vehicle that uses a combination of sensors, cameras, radar, and artificial intelligence (AI) to travel between destinations without a human operator. One of the features of such an advanced car is driver drowsiness detection. Driver drowsiness detection is a car safety technology that helps prevent accidents caused by the driver getting drowsy. Various studies have suggested that around 20% of all road accidents are fatigue-related, up to 50% on certain roads. There are various methods that can be used to find the drowsiness of the driver such as Steering pattern monitoring, Vehicle position in lane monitoring, Driver eye/face monitoring, etc.

Here, in this project students need to use the driver face monitoring technique to find the drowsiness.

This dataset contains images with the various position of the eyes i.e., Closed, Open, Yawn, No Yawn, etc. Students need to identify the class from the folder name of the dataset.

Project Outcomes:

- Explore various ways of computer vision methods to predict drowsiness.
- Identify various ML and DL approaches to predict drowsiness.
- Identify the best DL model to train the data.
- Train deep learning model for drowsiness.

8. **Project Title:** Aerial Perspective Object Detection

Source: [Semantic Drone Dataset](#)

Description:

Drone and aerial picture-taking quality has improved drastically in the past decade. Drone stabilization allows pictures taken from an aerial view to be crystal clear without shaking or blurriness. This has many practical and exciting applications for photography, cinematography, and also image recognition! Drone images can be used to quickly identify people and seek out specific objects in a large area. Think of how this could be used for spotting survivor rescues in disaster-struck areas.

The Semantic Drone Dataset focuses on semantic understanding of urban scenes for increasing the safety of autonomous drone flight and landing procedures. The imagery depicts more than 20 houses from nadir (bird's eye) view acquired at an altitude of 5 to 30 meters above the ground. A high-resolution camera was used to obtain images at a size of 6000x4000px (24Mpx).

The dataset consists of 400 images that have been annotated according to twenty standard classes such as trees, persons, cars, and pavement.

See more information at the dataset [website](#).

Project Outcomes:

- Identify everyday objects such as cars and roads in a bird's eye view images.
- Use a trained model to identify objects over a large, continuous, mapped area (i.e., your local neighbourhood from google maps).
- Use the positioning of cars and people determined to flag areas where pedestrians may be at most risk of an accident.

9. Project Title: Water Quality Detection

Source: [Water Quality Dataset](#)

Description:

Access to safe drinking-water is essential to health, a basic human right and a component of effective policy for health protection. This is important as a health and development issue at a national, regional, and local level. In some regions, it has been shown that investments in water supply and sanitation can yield a net economic benefit, since the reductions in adverse health effects and health care costs outweigh the costs of undertaking the interventions. In this project, students need to leverage this water quality dataset and apply machine learning and deep learning to distinguish potable and non-potable water.

There are 29 columns in the dataset. Some columns are:

Column	Description
pH	pH of water (0 to 14)
Hardness	Capacity of water to precipitate soap in mg/L
Solids	Total dissolved solids in ppm
Chloramines	Amount of Chloramines in ppm
Sulfate	Amount of Sulfates dissolved in mg/L
Conductivity	Electrical conductivity of water in $\mu\text{S}/\text{cm}$
Organic_carbon	Amount of organic carbon in ppm
Trihalomethanes	Amount of Trihalomethanes in $\mu\text{g}/\text{L}$
Turbidity	Measure of light emitting property of water in NTU
Potability	Indicates if water is safe for human consumption. Potable – 1 and Not Potable – 0

In this project students need to predict whether the water is potable or not by using water data.

Project Outcomes:

- Load and analyze the dataset.
- Visualize the dataset and data distribution.
- Pre-processing of the dataset.
- Build and design different machines learning models and FNN model.
- Use all features to train models and predict whether the water is potable or not.
- Find out using which features / data combination can get the best water quality detection model.

10. **Project Title:** Anomaly Detection in a Data Center with IoT Sensors

Source: [Dataset for Anomaly Detection in IoT Networks](#)

Description:

The rapid eruption of new interconnected environments – smart homes and cities, cyber-physical systems, health systems, etc. – has caused the emergence of the Internet of Things (IoT). The Internet and its users are under different kinds of threats. IoT environments are vulnerable to different kinds of anomalies and attacks that may cause unauthorized tasks to be performed by the remote malicious users. Thus, hackers, malicious software and viruses can disrupt the working of these devices. In one scenario an initial software failure due to an incorrect sensor configuration, caused temperature sensors to send incorrect data to the colling systems. The temperature of the devices was affected, causing a hardware failure, and the entire system collapsed.

This project is aimed at using machine learning and deep learning methods to detect real-world communication traffic anomalies using the Dataset for Anomaly Detection (DAD). The students in this project will perform an IoT data analysis to understand the normal and abnormal traffic behaviour of the network.

DAD is a semi-synthetic and labeled dataset consisting of seven days network activity with attacks spread over five days. It has three different types of anomalies: duplication, interception and modification on the MQTT message.

The anomaly traffic is made from IP source 10.6.56.41. The attacks have been done in specific days of the week. The distribution is as follows:

Monday 21st	there is no attack
Tuesday 22nd	some packets have been removed, so packets are not labeled as attack
Wednesday 23rd	a modification of packets is made between 4h and 6h
Thursday 24th	insertion of packets in less than 5 minutes at 3h
Friday 25th	a mix of interception, duplication and modification is done at 6h and between the 14-16h.
Saturday 26th	a mix of interception, duplication and modification is done at 6h and between the 14-16h
Sunday 27th	there is no attack

For more information check this [link](#).

Project Outcomes:

- Load and analyze the dataset.
- Visualize the dataset and data distribution.
- Pre-processing of the dataset.
- Build and design different machines learning models and FNN model to detect different types of anomalies.

11. Project Title: Energy Consumption of Household Appliances

Source: [Smart Home Dataset with Weather Information](#)

Description:

This project is similar to one of the earlier projects given in the list, but with a different dataset. Experimental data used to create regression models of appliances energy use in a low energy building.

The data set is at 1 min for about 350 days. The house temperature and humidity conditions were monitored with a smart meter.

There are 32 columns in the dataset. Some columns are:

Column	Description	Units
Time	time year-month-day	hour:minute:second
House overall	Overall house energy consumption	kW
Dishwasher	Energy consumed by the dishwasher	kW
Furnace 1	Energy consumed by the furnace 1	kW
Furnace 2	Energy consumed by the furnace 1	kW
Home office	Energy consumed by the home office	kW
Fridge	Energy consumed by the fridge	mm Hg

In this project students need to predict the usage of energy according to sensor collected data.

Project Outcomes:

- Load and analyze the dataset, visualize the dataset and data distribution.
- Pre-processing the dataset. Build and design different machine learning models and FNN model, if time permits, try RNN model.
- Find out using which features / data combination can get the best prediction for the energy consumption for some specific appliances.

12. Project Title: Smart Parking System

Source: [Dragon Lake Parking \(DLP\) Dataset](#)

Website Link: [Link](#)

Description:

The Smart Parking System project aims to develop an intelligent parking solution using machine learning and deep learning techniques. The project utilizes the Daron Lake Parking Dataset, which contains data on parking lot occupancy, vehicle counts, and other relevant information. By leveraging this dataset, the project will build a system that can accurately predict parking availability in real-time, enabling users to find parking spots more efficiently. The project will involve the following steps:

1. **Data Collection and Preprocessing:** The Daron Lake Parking Dataset will be acquired and preprocessed to remove inconsistencies, handle missing values, and ensure data quality.
2. **Feature Engineering:** Relevant features such as date, time, weather conditions, and special events will be extracted from the dataset to enhance the prediction accuracy.
3. **Model Development:** Machine learning and deep learning algorithms will be employed to create predictive models based on the dataset. Techniques such as regression, classification, and time series analysis may be utilized.
4. **Training and Evaluation:** The developed models will be trained using a portion of the dataset and evaluated using appropriate performance metrics such as accuracy, precision, recall, and F1-score.
5. **Real-Time Prediction:** The trained model will be integrated into a real-time system that takes input from various sources, such as sensors or cameras, to predict parking availability for different time intervals.
6. **Visualization and User Interface:** A user-friendly interface will be designed to display parking availability information to users, allowing them to make informed decisions about parking.

The Dragon Lake Parking (DLP) Dataset contains annotated video and data of vehicles, cyclists, and pedestrians inside a parking lot. Abundant vehicle parking maneuvers and interactions are recorded. To the best of our knowledge, this is the first and largest public dataset designated for the parking scenario (up to April 2022), featuring high data accuracy and a rich variety of realistic human driving behavior. To download trial sample or request full access, please visit the dataset webpage for more information.

Project Outcomes:

1. **Accurate Parking Prediction:** The project will deliver a machine learning model capable of accurately predicting parking availability based on historical data and various factors.
2. **Real-Time Availability Updates:** Users will have access to up-to-date parking availability information, ensuring they can find parking spots quickly and efficiently.
3. **Optimal Resource Utilization:** The Smart Parking System will aid in maximizing the utilization of parking spaces, reducing congestion, and improving overall efficiency.
4. **Improved User Experience:** With real-time information at their disposal, users can make informed decisions and save time searching for parking.
5. **Scalability and Adaptability:** The project will consider the scalability and adaptability of the system to accommodate additional parking lots and handle various types of data sources.
6. **Insights and Analytics:** The analysis of the dataset and model performance can provide valuable insights into parking patterns, peak hours, and factors influencing parking availability.

13. **Project Title:** Indoor Localization and Navigation System

Source: [Microsoft Indoor Localization Dataset](#)

Website Link: [Link](#)

Description:

The Indoor Localization and Navigation System project aims to develop a robust and accurate solution for indoor positioning and navigation using machine learning techniques. Leveraging the Microsoft Indoor Localization Dataset, which provides sensor data from WiFi, Bluetooth, and inertial sensors, this project will focus on creating a system that can accurately determine the location of individuals within indoor environments and provide navigation guidance.

The project will involve the following steps:

1. **Dataset Exploration:** The Microsoft Indoor Localization Dataset will be analyzed to understand its structure and the information it contains, including WiFi and Bluetooth signal strengths, sensor readings, and ground truth locations.
2. **Feature Extraction:** Relevant features will be extracted from the dataset, such as WiFi and Bluetooth signal strengths, sensor data (e.g., accelerometer, gyroscope), and other contextual information available, to improve localization accuracy.
3. **Model Development:** Machine learning algorithms, such as k-nearest neighbors, support vector machines, or deep learning models, will be employed to create a localization model based on the dataset. The model will be trained to predict the location of an individual based on the provided sensor data.
4. **Localization and Navigation:** The developed model will be integrated into a real-time system that takes sensor data from a user's device and provides their current indoor location. The system will further offer navigation guidance to reach a target location within the indoor environment.
5. **Evaluation and Optimization:** The system's accuracy and performance will be evaluated using appropriate metrics, such as mean localization error or precision-recall curves. Techniques like parameter tuning and model optimization will be explored to enhance the system's accuracy.

Project Outcomes:

1. **Accurate Indoor Localization:** The project will deliver a machine learning-based system capable of accurately determining a user's indoor location using the Microsoft Indoor Localization Dataset.
2. **Real-Time Positioning:** The system will provide real-time updates of the user's location, enabling seamless indoor navigation.
3. **Navigation Guidance:** Users will receive step-by-step navigation instructions to reach their desired locations within the indoor environment.
4. **Adaptability to Various Indoor Environments:** The system will be designed to work across different indoor environments, accommodating variations in building structures, layouts, and signal characteristics.
5. **Sensor Fusion:** The project will explore techniques to combine data from multiple sensors, such as WiFi, Bluetooth, and inertial sensors, to improve localization accuracy.
6. **Optimal Path Planning:** The navigation system will compute optimal paths considering factors like distance, obstacles, and user preferences.
7. **User Experience Enhancement:** The project aims to improve the overall indoor navigation experience, reducing frustration and time spent searching for specific locations.
8. **Insights and Analytics:** The analysis of the Microsoft Indoor Localization Dataset and system performance can provide insights into indoor mobility patterns, occupancy levels, and factors influencing localization accuracy.

14. **Project Title:** Intelligent Fire Detection and Alert System

Source: [Fire Detection Dataset](#)

Description:

The Intelligent Fire Detection and Alert System project aims to develop a machine learning-based solution for early fire detection and timely alerts. Utilizing the Fire Detection Dataset from Kaggle, which consists of images labeled with fire and non-fire classes, this project focuses on creating a system that can accurately identify and detect fires in images or video streams..

The project will involve the following steps:

1. **Dataset Exploration:** The Fire Detection Dataset will be analyzed to understand the structure, size, and distribution of fire and non-fire images. Exploratory data analysis will help gain insights into the characteristics of fire-related images.
2. **Data Preprocessing:** The dataset will undergo preprocessing steps such as resizing, normalization, and augmentation to prepare it for training machine learning models. This step may involve techniques like image cropping, rotation, and flipping.
3. **Model Development:** Machine learning algorithms, such as convolutional neural networks (CNNs), will be employed to train a fire detection model based on the dataset. The model will learn to classify images as fire or non-fire.
4. **Model Training and Evaluation:** The developed model will be trained using a portion of the dataset and evaluated using appropriate metrics such as accuracy, precision, recall, and F1-score. Techniques like cross-validation may be used to assess the model's performance.
5. **Real-Time Fire Detection:** The trained model will be integrated into a real-time system that takes input from images or video streams, and accurately identifies and detects fires.
6. **Alert System:** Upon fire detection, the system will generate immediate alerts, which can include visual notifications, audible alarms, and notifications sent to designated individuals or emergency services.
7. **Performance Optimization:** Techniques such as hyperparameter tuning, model ensembling, or transfer learning will be explored to enhance the model's accuracy and robustness.

Project Outcomes:

1. **Accurate Fire Detection:** The project will deliver a machine learning model capable of accurately detecting fires in images or video streams using the Fire Detection Dataset.
2. **Real-Time Fire Alerts:** The system will provide immediate alerts upon fire detection, enabling timely response and reducing potential damage.
3. **Early Fire Detection:** By detecting fires at an early stage, the system can help prevent fire incidents from escalating and spreading.
4. **Robust Performance:** The system will be optimized to handle various environmental conditions, lighting variations, and different types of fire instances.
5. **Minimized False Positives:** Efforts will be made to reduce false positives, ensuring that the system accurately distinguishes between actual fires and other non-fire elements.
6. **Scalability and Adaptability:** The system will be designed to scale and adapt to different camera sources, including CCTV cameras, surveillance systems, or image/video feeds from multiple devices.
7. **Real-World Deployment:** The project will consider the practical implementation and deployment of the fire detection system in real-world settings, such as residential or commercial buildings.
8. **Insights and Analytics:** The analysis of the Fire Detection Dataset and system performance can provide insights into fire patterns, factors influencing detection accuracy, and contribute to fire safety research.

15. **Project Title:** Occupancy Detection in Smart Buildings

Source: [UCI Occupancy Detection Dataset](#)

Website Link: [Link](#)

Description:

The Occupancy Detection in Smart Buildings project aims to develop a machine learning solution for accurately detecting occupancy in smart buildings using sensor data. Leveraging the UCI Occupancy Detection Dataset, which contains environmental sensor readings such as temperature, humidity, CO2 levels, and light intensity, this project focuses on creating a system that can predict occupancy status in real-time. The project will involve the following steps:

1. **Dataset Exploration:** The UCI Occupancy Detection Dataset will be analyzed to understand its structure, size, and the relationship between sensor readings and occupancy status. Exploratory data analysis will help gain insights into the data characteristics.
2. **Data Preprocessing:** The dataset will undergo preprocessing steps, including handling missing values, feature scaling, and feature engineering. This step may involve techniques such as normalization, outlier detection, and handling categorical variables.
3. **Model Development:** Machine learning algorithms, such as logistic regression, decision trees, random forests, or neural networks, will be employed to train a model based on the dataset. The model will learn to predict occupancy status based on the provided sensor data.
4. **Model Training and Evaluation:** The developed model will be trained using a portion of the dataset and evaluated using appropriate metrics such as accuracy, precision, recall, and F1-score. Techniques like cross-validation may be used to assess the model's performance.
5. **Real-Time Occupancy Detection:** The trained model will be integrated into a real-time system that takes input from the building's sensors and provides occupancy predictions in real-time.
6. **Energy Optimization:** By accurately detecting occupancy, the system can contribute to optimizing energy consumption in smart buildings by adjusting heating, cooling, lighting, and ventilation systems accordingly.
7. **Alert System:** The system can trigger alerts or notifications based on occupancy status, allowing facility managers or occupants to monitor and respond to occupancy changes.

Project Outcomes:

1. **Accurate Occupancy Detection:** The project will deliver a machine learning model capable of accurately predicting occupancy status based on sensor data from the UCI Occupancy Detection Dataset.
2. **Real-Time Occupancy Updates:** The system will provide real-time occupancy updates, allowing building occupants or facility managers to monitor occupancy changes.
3. **Energy Efficiency:** By optimizing energy consumption based on occupancy, the system can contribute to reducing energy waste and improving the overall energy efficiency of smart buildings.
4. **Cost Savings:** Efficient occupancy detection can lead to cost savings by optimizing energy usage and reducing unnecessary maintenance or operations.
5. **User Comfort:** The project will consider the impact of occupancy on user comfort, ensuring that the system supports a comfortable indoor environment for building occupants.
6. **Scalability:** The system will be designed to scale and adapt to different smart building environments, accommodating varying sensor configurations and building layouts.
7. **Integration with Building Management Systems:** The project will explore the integration of the occupancy detection system with existing building management systems or automation systems, enhancing overall building operations and control.

16. **Project Title:** Occupancy Detection in Smart Buildings

Source: [UCI Occupancy Detection Dataset](#)

Website Link: [Link](#)

Description:

The Occupancy Detection in Smart Buildings project aims to develop a machine learning solution for accurately detecting occupancy in smart buildings using sensor data. Leveraging the UCI Occupancy Detection Dataset, which contains environmental sensor readings such as temperature, humidity, CO2 levels, and light intensity, this project focuses on creating a system that can predict occupancy status in real-time. The project will involve the following steps:

1. **Dataset Exploration:** The UCI Occupancy Detection Dataset will be analyzed to understand its structure, size, and the relationship between sensor readings and occupancy status. Exploratory data analysis will help gain insights into the data characteristics.
2. **Data Preprocessing:** The dataset will undergo preprocessing steps, including handling missing values, feature scaling, and feature engineering. This step may involve techniques such as normalization, outlier detection, and handling categorical variables.
3. **Model Development:** Machine learning algorithms, such as logistic regression, decision trees, random forests, or neural networks, will be employed to train a model based on the dataset. The model will learn to predict occupancy status based on the provided sensor data.
4. **Model Training and Evaluation:** The developed model will be trained using a portion of the dataset and evaluated using appropriate metrics such as accuracy, precision, recall, and F1-score. Techniques like cross-validation may be used to assess the model's performance.
5. **Real-Time Occupancy Detection:** The trained model will be integrated into a real-time system that takes input from the building's sensors and provides occupancy predictions in real-time.
6. **Energy Optimization:** By accurately detecting occupancy, the system can contribute to optimizing energy consumption in smart buildings by adjusting heating, cooling, lighting, and ventilation systems accordingly.
7. **Alert System:** The system can trigger alerts or notifications based on occupancy status, allowing facility managers or occupants to monitor and respond to occupancy changes.

Project Outcomes:

1. **Accurate Occupancy Detection:** The project will deliver a machine learning model capable of accurately predicting occupancy status based on sensor data from the UCI Occupancy Detection Dataset.
2. **Real-Time Occupancy Updates:** The system will provide real-time occupancy updates, allowing building occupants or facility managers to monitor occupancy changes.
3. **Energy Efficiency:** By optimizing energy consumption based on occupancy, the system can contribute to reducing energy waste and improving the overall energy efficiency of smart buildings.
4. **Cost Savings:** Efficient occupancy detection can lead to cost savings by optimizing energy usage and reducing unnecessary maintenance or operations.
5. **User Comfort:** The project will consider the impact of occupancy on user comfort, ensuring that the system supports a comfortable indoor environment for building occupants.
6. **Scalability:** The system will be designed to scale and adapt to different smart building environments, accommodating varying sensor configurations and building layouts.
7. **Integration with Building Management Systems:** The project will explore the integration of the occupancy detection system with existing building management systems or automation systems, enhancing overall building operations and control.

