**DOCUMENTATION**

(according to everyday learning for the given tasks)

DAY 1:

1. LiDAR:

* It stands for Light detection and ranging which is a remote sensing method.
* Light is sent from a source and reflected from objects in the scene.
* It is used to measure scatter, absorption or re emission from particles or molecules in the atmosphere.
* It is used to determine the velocity of the target.

1. Software used: The following softwares are used:

* Code V: It is a optical design software for designing receiver optics in a LiDAR system.
* Light Tools: It is used for modelling and analyzing LiDAR.

1. Working: It emits waves of energy to detect and track objects which makes use of reflected light which can measure distance faster with higher precision and resolution.

4.File Format: There are two different file formats for the dataset:

* .las: It stores LiDAR point cloud data
  + - * 1. It is supported by American Society of Photogrammetry & Remote Sensing.
* .laz: It was developed by Martin Isenberg of LasTools and is highly compressed version.

1. Libraries: The libraries which are used are – libLas and .pyLAS . They both provide only read and write access to LAS data.
2. Tools: The tools which is used is FUSION LTK.

6.Steps in Preprocessing: The different followed while preprocessing the las/laz dataset are:

* QUALITY CHECK: It involves identification and rectification of issues such as data gaps, intensity anomalies.
* REMOVE NOISE & ARTIFACTS: Such as vegetations and buildings are removed.
* CLASSIFY THE DATA: It classifies the data based on different characteristics which includes ground points, vegetation points, building points and many more.
* CORRECT FOR SYSTEMATIC ERRORS: It includes correction of error in sensor calibration. The correction techniques include- adjusting elevation values and performing atmospheric correction.

8.Types of Classification: The different types of classification include:

* AUTOMATIC: It makes uses of machine learning algorithm to classify the points into buildings, ground etc.
* MANUAL: It is performed when super precision is required.
* FEATURE: It is performed for finding and pulling out the specific features.
* CUSTOM: it is performed as per the user requirements.
* QUALITY CONTROL: It makes sure that data is accurate and reliable.

DAY 2:

1.Features of .las File: The different features of las files are as below:

* XYZ COORDINATES: They basically are the coordinates of each point in which ‘X’ represents the horizontal axis, ‘Y’ represents the vertical axis and ‘Z’ represents the elevation.
* INTENSITY: Represents the strength of the laser return signal from the object surface.
* RETURN NUMBER: It is the order of the point in the sequence of returns from a single laser.
* NUMBER OF RETURN: It is total number of returns generated by the a single laser pulse.
* SCAN DIRECTION FLAG: Indicates the direction of the laser scan where-

0: Left to right.

1: Right to left.

* EDGE OF FLIGHT LINE: A flag (0 or 1) that identifies whether the point lies at the edge of the flight line.
* CLASSIFICATION: Categorizes each point based on its type or surface The different classification codes are:

1. 0 : which are never classified
2. 1: Unclassified.
3. 2: Ground.
4. 3: Low Vegetation.
5. 4: Medium Vegetation.
6. 5: High Vegetation.
7. 6: Building.
8. 7: Noise.
9. 8: reserved
10. 9: Water.
11. 10: Bridge Deck

* POINT SOURCE ID: Identifies the source flight line or segment that generated the point.
* GPS SYSTEMS: Records the precise time (in GPS seconds) when the laser pulse was emitted.

2.Point Cloud:

* Point cloud refers to the collection of data points in a three-dimensional space that represent the surfaces of objects captured by the LiDAR sensor.
* When visualized, the points collectively resemble a cloud. They are scattered across space, representing the surfaces of physical objects (such as terrain, buildings, or trees), forming a dense "cloud" of points.
* They consist only of discrete points which means that the dataset represents the scanned area as a collection of individual data points in 3D space, rather than as a continuous surface or object.

Day 3:

1.CRS systems:

* It stands for Cordinate Reference System.
* They are systems used to define how spatial data is represented in a 3D space.
* It is important for understanding and working with geospatial data, as they define the relationship between real-world locations and their coordinates in the datasets.
* It has different components:

1. Coordinate Systems which define how positions are measured. The common type used is cartesian.
2. Datum is a reference model for Earth’s shape and size
3. Projection converts the 3D coordinates into 2D cartesian coordinates for mapping.
4. Units specifies that what unit CRS should use whether it be feet, degrees or metres.

* It has 3 types as follows:

1. GEOGRAPHICAL CRS: It is CRS that uses a 3D spherical representation of Earth (latitude and longitude) based on an angular coordinate system. The coordinates are measured in degrees and it is used commonly for navigation and positioning.
2. PROJECTED CRS: It projects the 3D surface of Earth onto a 2D plane using a mathematical projection. The coordinates are expressed in either feet or meters.
3. VERTICAL CRS: A CRS that specifies heights or depths relative to a vertical reference.

2.PDAL:

* It is Point Data Abstraction Library.
* It is a C++ library which provides robust tools to manipulate, filter, transform and analyze point cloud data.
* It works with LAS, LAZ as well as CSV.
* It’s one of the important feature is pipeline processing which enables to define a sequence of operations (e.g., filter, classify, export) in JSON or command-line format.

3.Return number VS Number of Return:

* Return Number: It helps distinguish between different layers of objects.

Ex: when a laser pulse hits multiple surfaces, the sensor records multiple returns for that pulse as

1: The first object hit (e.g., the top of the tree which is canopy)

2: The second object hit (e.g., branches).

3: The last object hit (e.g., the ground).

* Number of Returns: The number of returns specifies the total number of returns detected for a single laser pulse.

Example: A laser pulse hits the flat ground surface it can have Number of Returns = 1.

* A laser pulse passing through vegetation may have Number of Returns = 3 (one for the canopy, one for a branch, and one for the ground).

DAY 4:

1.Incremental Learning:

* In incremental learning, model is trained continuously.
* It never forgets previously acquired information.
* It operates with limited memory, making it suitable for scenarios where storing all past data is infeasible.
* It handles environments where data distribution changes over time.
* It reduces computational costs by focusing only on new data or changes in the dataset.
* It first trains the model on the base dataset which is provided and if any modifications is made to the old dataset or when the new dataset is available, it immediately trains the model and make the adjustments required and periodically assess the model’s performance to make sure that both the dataset is handled effectively.

2.River ML:

* It is a Python library for building online machine learningsystems basically incremental learning.
* In river ML, the data arrives continuously and model updates immediately.
* It supports handling concept drift which is one of the challenges of incremental learning.
* It works seamlessly with libraries like pandas, NumPy, and scikit-learn.
* It includes algorithms like regression( linear and logistic), K-means clustering and classification.
* It also helps in detecting the concept drift.
* River ML can be used as it provides different features like- quick adaptation, scaling large dataset, real time updation.

3.Deep River ML:

* It is combination of deep learning principles with the streaming and incremental learning approaches.
* It serves as extension of river

4.Model drift:

* Model drift refers to the phenomenon that occurs when the performance of a machine learning model degrades with time. It is of 2 types-

1. CONCEPT DRIFT:

* It occurs when there is a change in the functional relationship between a model’s input and output data.
* Also known as model drift.

1. DATA DRIFT:

* It occurs when the distribution of the input data changes over time.

The solution to model drift is Kolmogorov Smirnov test which is a statistical test which is used to determine where the two sets of data come from same distribution or not. It is of 2 type:

One sample test

Two sample test

They make use of Commulative Distribution Function and Empirical commulative distribution Function.

6.Importance of Skewness, Kurtosis, Variance, Standard deviation:

1. SKEWNESS: It measures the asymmetricity of the data distribution. The importance of it includes:

* Height distribution indicates whether the elevation values are symmetrically distributed or biased towards the higher and lower values.

Example: In vegetation analysis, positive skewness suggests a larger number of lower elevation points which is basically ground and fewer tall objects like trees.

* Extreme skewness can signal the presence of outliers

1. KURTOSIS: It measures the degree of peakness or flatness of the graph.

* It helps in distinguishing between smooth and irregular surfaces.
* A high kurtosis value could highlight unusual structures like towers.
* Low kurtosis indicates a relatively flat terrain.

1. VARIANCE: Variance measures the spread or dispersion of the dataset around its mean.

* High variance in elevation data suggests diverse terrain (e.g., mountainous areas), while low variance indicates flat surfaces (e.g., plains).
* Variance across tiles or patches helps identify inconsistencies in data collection.

1. STANDARD DEVIATION: Standard deviation is the square root of variance which provides more detailed measure of spread of the data from its mean value.

* A high standard deviation in flat regions might indicate noise or errors in data collection.
* Consistent standard deviation across tiles ensures uniformity in dataset.

7. Patches & Tiles:

* Tiles are rectangular or square sections of the LiDAR dataset which is created by splitting the point cloud into grid based structure.
* Each tile covers a geographical area and contains all the points of that area
* It helps in managing the large dataset as it breaks them into smaller chunks.
* Patches refers to smaller subset of points created from a single tile.
* They are used for detailed analysis.
* It focuses on specific area within the tile for finer analysis.
* A tile is divided inti patches for the ground classification, object detection or when there is need to make clusters out of it.