Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Computer Science and Engineering (Data Science)

Name: Ananya Godse **SAP ID:** 60009220161 **Batch:** D1 – 2

Solar Power Generation Forecasting

1. Describe your problem in detail and discuss why it is a data science problem.

   Our world is on the brink of a climate crisis, driven primarily by the accumulation of greenhouse gases in the Earth's atmosphere. These greenhouse gases are released when fossil fuels are burned. According to the Government of India's NITI Aayog website, in 2022, 58.63% of our energy supply came from the burning of coal and 29.32% from oil. That means that close to 88% of our energy supply comes from non-renewable, climate change causing sources. Clearly, renewable sources of energy are the need of the hour.

   Fortunately, India is making strides in this area. One such source of renewable energy is solar energy. Solar Power grids are being laid down every day, increasing our power generation capacity. But there is an inherent variability to the production of solar energy. Its dependent on weather conditions, time of the day, seasonal changes, and geographic factors.

   If we are going to rely on solar energy to fulfil a larger slice of energy consumption, we need to ensure that it will be enough. Solar Power Generation Forecasting is thus necessary to manage the logistics of electricity supply and optimize grid management.

   The problem here is to predict how much power a solar power plant will generate on any given day based on the weather.

   This is a data science problem because it involves analysing large volumes of data from various sources (weather forecasts, historical energy production data, geographical information, etc.) to build accurate predictive models that can anticipate fluctuations in solar energy production. These models are crucial for optimizing the efficiency and reliability of solar energy systems and integrating them effectively into the broader energy infrastructure.

2. Justify that the data chosen is appropriate to build a model to solve the problem.

   Dataset Link: https://www.kaggle.com/datasets/anikannal/solar-power-generation-data?resource=download&select=Plant_1_Generation_Data.csv

   This data has been gathered at two solar power plants in India over a 34-day period. It has two pairs of files - each pair has one power generation dataset and one sensor readings dataset. The power generation datasets are gathered at the inverter level - each inverter has multiple lines of solar panels attached to it. The sensor data is gathered at a plant level - single array of sensors optimally placed at the plant.

   Since this is data is collected from a solar power plant in India and there is data about the solar energy yield and data from weather sensors, this dataset is perfect for figuring out how much solar energy will be produced on any given day based on weather factors.