**Initiative of NSF (National Science Foundation)**

**Harnessing the Data Revolution (HDR): Institutes for Data-Intensive Research in Science and Engineering - Ideas Labs (I-DIRSE-IL)**

**Data-Intensive Research in Science and Engineering (DIRSE)**

There are numerous science and engineering challenges that require, or will soon require, data science to help address research and technological questions. Some examples include: near-term ecological forecasting; understanding how the phenotype of living things is determined by their genotype and environment; real-time sensing, learning, and decision making for resilient engineering systems; development of autonomous technologies; predictive understanding of the Earth system which includes climate, weather, hydrologic, seismic, and space weather hazards; multi-messenger astrophysics; understanding the nature of dark matter; predictive design of next-generation catalysts; elucidation of design rules for emergent molecular properties from atomic-scale interactions; design of sustainable chemical manufacturing systems; real-time optimization and control of complex chemical and biological systems; discovery of new advanced materials; integration of heterogeneous data for explaining human behavior, learning, and social processes; understanding the brain, including prediction of complex systems for neuroimaging and neurological applications; and understanding student learning and success across STEM disciplines. Advancing knowledge in these areas requires solutions to many modeling and data challenges such as real-time sensing, learning, and decision making; social, political, and behavioral implications of machine learning and impacts of new data uses; issues related to ethics and fairness; and integrating heterogeneous data for explaining or predicting complex phenomena. There is also a need for approaches that combine physical models with data driven models for learning and decision making.

Data science tools, such as signal and image processing, visualization, statistical modeling and inference, machine learning, and optimization, offer a starting point for solving important scientific and engineering challenges. However, extracting new information and knowledge from data will benefit from new, convergent strategies that capitalize on existing NSF investments in data and cyberinfrastructure and that build synergy between the researchers with expertise in the generation or measurement of data and those with expertise in processing and analyzing that data.

**Recommendation system for Irrigation:**

(Real-time sensing, learning, and decision making for resilient engineering systems)

Large amount of water in irrigated agriculture is wasted due to poor water management practices. To improve water management in irrigated areas, models for estimation of future water requirements are needed. In this study, we prepare a data set containing information on suitable attributes. The data are obtained from three different sources namely water delivery statements, meteorological data, and remote sensing images. In order to make the prepared data set useful for demand forecasting and pattern extraction the datasets are to be processed using a novel approach based on a combination of irrigation and data science knowledge.

This will require pre-processing of data before any models can be applied as the data is sourced from several disparate systems

One of the key parameter will be to get the real time data about the soil moisture before the crops are irrigated. This data is to be combined with other data that is acquired from other disparate systems mentioned in the above paragraphs. New models need to be defined considering all these data sources in order to provide a right recommendation to the farmers. This helps in the productivity and also to maintain the ecology in a certain way.