my projects:
anomaly detection in time series of traffic data
influencer detection in call/sms data (CDR) by graph creation and information spread simulations
website classification and content analysis
analysis of DNS data, cleaning dns data from ads and automatically generated content

Rave - software that has data coming from different sources in one place, makes clinical trials easier
easier access to real-time patient data
and data monitoring
using machine learning tools to detect discrepancies, anomalies, data tendencies, making predictions
visualization
making decisions

MEDS development:

MEDS isn't simply a place where all data is stored or co-located. It's the platform where data can be transformed, extracted, and accessed via a suite of robust tools and solutions.
Unlike a traditional data warehouse, MEDS provides the orchestration and unification to create a single source of truth for all clinical data captured or ingested in the Medidata Rave Clinical Cloud™ platform.

CDS (clinical decision support) workbench development

1. Is the patient a candidate for the drug, and will the patient likely respond to the drug?
2. What is the response the drug gets after the market release, and how can I select/change therapy based on this information?
3. How should the drug be administered, including dosage calculations? (say appropriate for people having certain deceases, etc.)

A **genome** is an organism's complete set of DNA, including all of its genes. Each **genome** contains all of the information needed to build and maintain that organism.
Genomics - applying the techniques of genetics and molecular biology to the genetic mapping and DNA sequencing of sets of genes or the complete genomes of selected organisms, with organizing the results in databases, and with applications of the data
Genetic mapping offers evidence that a disease transmitted from parent to child is linked to one or more genes and provides clues about which chromosome contains the gene and precisely where the gene lies on that chromosome.

DNA sequencing - determine the order of the four bases: adenine, guanine, cytosine, and thymin

Genomic data + trial data = will improve the quality of trials
genomic data improves inclusion/exclusion criteria, adaptive interventions, targeted therapies, etc.

Data mining Algorithms for **Healthcare Decision Support System**

decision tree
naive bayes (simple probabilistic classifier)

NN

Data mining in Clinical trials

Tracking patterns: Tracking patterns is one of the simple techniques used to understand and find out patterns in the concerned dataset. This is usually recognition of the unusualness of data occurring at regular intervals.
Classification: collection of various aspects together into obvious categories, which can be further used to draw conclusions.
Association: Association is related to following patterns but is more specific to **dependently related variables.** In this type of technique, specific events or attributes which are extremely connected with another event are looked into.
Outlier detection: In this type of technique, we don't only look into patterns, but we also look into the outliers in the data.
Clustering: Clustering is like classification. However, it involves grouping portions of data organised collectively **on the basis of their similarities**.
Regression: It is a form of arranging and demonstrating the possibility of certain variables in the presence of other variables.
Prediction: This is one of the most treasured data mining techniques. It plans for the type of data which will be seen in future, just by recognising the patterns of the past, which are enough to chart a precise forecast.

as simple as probabilistic models (naive bayes)