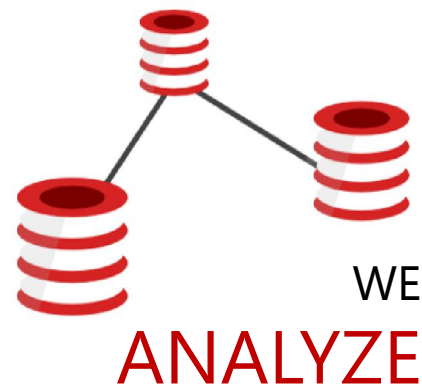
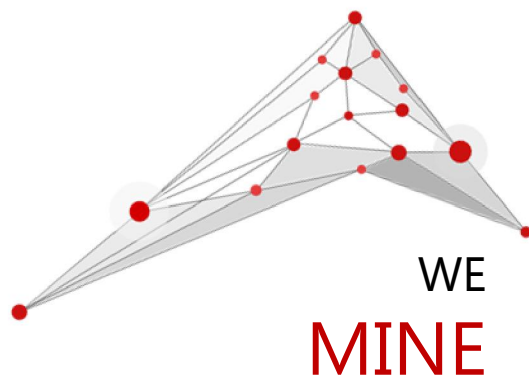


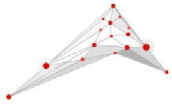


## DATA EXPERTS

We accelerate research and transform data to help you create actionable insights



# Domains



## Data Mining

Mining longitudinal and linked datasets from web and other archives.



## Data Standardization

Converting raw data into cleaned and standardized formats.



## Big Data Analytics

Implementing distributed computing methods and cutting edge statistical models to analyze huge chunks of data.



## Statistical Modeling

Curating statistical models or running statistical analysis on real or simulated data to understand phenomenon.



## Data Visualization

Creating customized visualization for your data-sets to tell a compelling story.



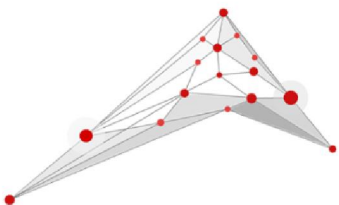
## Technology Implementations

Implementing customized mobile and software tools to make research faster, scalable and economical.

# Data Mining

Using our proprietary codes, we can deliver harvested data from web or documents in a longitudinal format.

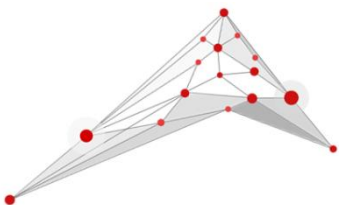
- ➔ Codes built on open source technologies which can run on most popular operating systems
- ➔ Quick turnaround time
- ➔ Automated quality layer can be embedded customized to the data-sets
- ➔ Logs are maintained throughout the lifecycle
- ➔ Harvested data is delivered in flat file formats usable by any database platform



# Data Mining – Indian Patent Data

Task in hand was to write a code to harvest Indian Patent Data from Indian Patent Office Website on an ongoing basis

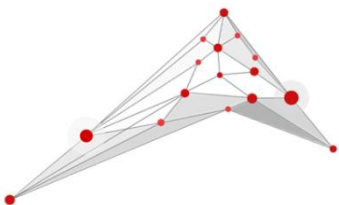
- ➔ Code was prepared to extract information on all patents filed between date A and date B defined by user.
- ➔ Further, a data quality tool was written to clean data from the harvested pool of data.
- ➔ A scheduling code was prepared to run harvesting code followed by data quality code every week (as Indian Patent office publishes patents every week).
- ➔ This master scheduler code runs every week and updates all fields of new patents in the master pool of database.



# Data Mining – Earning Conference Calls

Task in hand was to write a code to harvest latest earning conference calls of all NASDAQ listed firms and parse each call at individual speaker level

- ➔ Earning Conference calls were harvested from NASDAQ, Thomson StreetEvents and other sources on a daily basis
- ➔ Automation code was written to parse those conference calls at individual speaker level with meta data like speaker name, designation, company and speech text
- ➔ A scheduling code was prepared to run harvesting code and parsing code followed by harvesting every day (from NASDAQ)
- ➔ This master scheduler code runs every day and updates all fields of new conference calls in the master pool of database.



# Statistical Modeling

Our team of Statisticians make complicated statistical models from scratch and our coders and designers make them a usable reality

- ➔ Modeling carried out statisticians from top institutes like IITs and IISCs with experience in firms like Oracle, Global Analytics, and MuSigma
- ➔ Statistical Modeling is carried out on technologies like R, Matlab, GLPK, Gurobi
- ➔ Our coders make these models scalable and efficient
- ➔ UI/UX designers make them user-friendly and flexible as per working requirements
- ➔ Harvested data is delivered in flat file formats usable by any database platform



# Stats – School Cost Optimization

Task in hand was to prepare, implement, and realize statistical models to optimize school transportation and logistics costs for cities in a Latin American Country

- ➔ A Professor was appointed by NPO and a Latin American government to prepare statistical models aiming at city planning for optimizing school costs
- ➔ Our statisticians used clustering techniques, Ant Colony Optimization, and multi-processing to minimize total transportation + operational cost on schools for a city.
- ➔ Various constraints and difficulties were taken care of like bus capacity, total riding time, mixed loading, land and sea routes, different types of buses and classroom sizes, and many more.
- ➔ A graphical user interface with flexibility was delivered to client to view output on maps, spreadsheets, manually change routes, upload input data etc.



# Stats – HMM predicts App Lifetime

Using Hidden Markov Models to predict app lifecycle and various hidden layers of app usage by categories

- ➔ Using a hypothesis that users have different lifecycles in using category based apps, for eg. Productivity apps for first two months followed by entertainment in next two months, we will predict lifecycle of an app on an average user's device.
- ➔ Using data on installs and de-installs of many app for 100,000 customers, and modeling it on a Hidden Markov Model, we estimated lifecycles of 6 categories of apps on an average customer's device.
- ➔ Finally, these lifecycles were used to correlate life-cycle of individual apps using various app parameters.
- ➔ This analysis was order-complex and time consuming. Thus an Hadoop layer on R was used to reduce time and memory complexity on a single machine using HDFS.

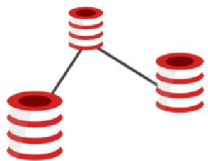




# Data Analysis (Big Data)

Data Transformation and Analytics at large scale to deliver insights

- ➔ Our team of coders and domain analysts work together to code transformation and analytical processes.
- ➔ These codes are made scalable, time efficient, and memory efficient when databases runs beyond megabytes of memory and millions of rows using Hadoop and Distributed File Systems.
- ➔ With knowledge and command on languages like Python, SQL, Hadoop, NoSQL, R, Pig, MapReduce, our coders can make any analysis happen fast enough and our domain analysts can deliver useful insights in best representations possible.



# Data Analysis – Sentiment Analysis

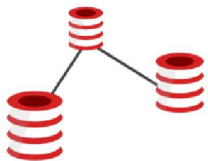
Navigating through millions of tweets to understand sentimental behavior of customers towards brands

➔ Millions of tweets talking about few famous brands were extracted using data harvesting techniques.

➔ Using Naïve Bayes Analysis and a training set of 100,000 tweets, a supervised classifier was established to decide sentiment of any tweet. This analysis involved prior cleaning of stop words and substituting prolonged chat versions of words with their English word counterparts.

➔ Sentiments of customers were analyzed for various brands. Further, change in sentiments of customers were analyzed after a brand re-tweets and engages with the customer to examine effects of customer engagement by big brands on twitter.

➔ This project was not feasible to complete in a record two month timeframe without use of big data technologies.



# Data Analysis – Personality Analysis

Personality of Chief Executives and Analysts during Conference Calls were examined and correlated with performance of firms in following quarter

- ➔ Having access to parsed database of various earning conference calls of NASDAQ traded firms, we performed personality analysis of over 30,000 people (including chief executives and analysts) using psycho-linguistic databases of words and SVM technique.
- ➔ Each person was given a score from 1 to 8 in big five personality traits.
- ➔ These traits were correlated with following quarters conference calls and patterns were observed in personality of chief executives and their effect on financial (especially stock) performance in the following quarter.



# Clients

Researchers from

HARVARD  
BUSINESS  
SCHOOL

STANFORD  
GRADUATE  
SCHOOL OF  
BUSINESS

FUQUA  
SCHOOL OF  
BUSINESS

WHARTON  
SCHOOL

MIT  
SLOAN

HAAS  
SCHOOL OF  
BUSINESS

LONDON  
BUSINESS  
SCHOOL

HEC  
PARIS

NYU STERN  
SCHOOL OF  
BUSINESS

COLUMBIA  
BUSINESS  
SCHOOL

NANYANG  
BUSINESS  
SCHOOL

NUS  
BUSINESS  
SCHOOL

INSEAD

IESE  
BUSINESS  
SCHOOL

TUCK  
BUSINESS  
SCHOOL

# Contact



**Innovaccer**

D-66

Sector 63

Noida, UP - 201301

India

+91 120 431 1139

Sachin Jaiswal

VP, Business Development

(P): +91 - 120 431 1139

(E): [sachin.jaiswal@innovaccer.com](mailto:sachin.jaiswal@innovaccer.com)

