DECiMAL POiNT

Innovative Data & Research Solutions

Data Science and Analytics

Challenges and Applications in Finance

November 2022

Key Trends in Data Science

Implosion in Data is One of the Key Recent Trends Witnessed Globally…

2.5 Quintillion

Data Generated in Bytes by Internet Users Globally Each Day

79 Zettabytes

Data Created, Consumed, and Stored in 2021

40%

Share of Machine-generated Data on the Internet in 2021

70%

of Asset Managers in the US Prefer Using Alternative Data

95.0%

Businesses Facing Unstructured Data Handling Challenges

59%

of Asset Managers Currently Factor ESG Data into their Investment Process

Data Overload with the Availability of Almost Too Much Data in Varied Forms

Data Democratization with Faster Dissemination of Data to More Players

Markets are Reacting Quicker and More Violently to New Data with the Spread of Trading Algorithms

This is Challenging to Sustain an Alpha Edge

…With Alternative Data in Prime Focus

Technology and Improved Connectivity has seen a Manifold Increase in Sources of Alternate Data

- **Credit and Debit Card Transactions**
- **Email Receipts**
- **Geo-location (Foot Traffic) Records**
- **Mobile Application Usage**
- **Satellite and Weather Data**
- **Social Platforms**

**Not a Comprehensive List*

BFSI Sector Leads in the Usage of Alternative Data; Other Early Adopters include Online Retailers, SaaS Purveyors and Hospitality

Organizations are Increasingly Marrying Alternative Data with Traditional Data for Risk Management

GDPR, CCPA, and other Privacy Regulations Pose Threats to some of the Alternative Data Sources

Organizations are Turning to AI and ML to Leverage Alternate Data for Generating Meaningful Insights

Advanced Analytics and Data Science are the Key to Examine and Understand Alternative Data Sets

Finance Companies are Increasingly Turning towards Technology, Especially AI and ML to Leverage this Data

Quick Ingestion

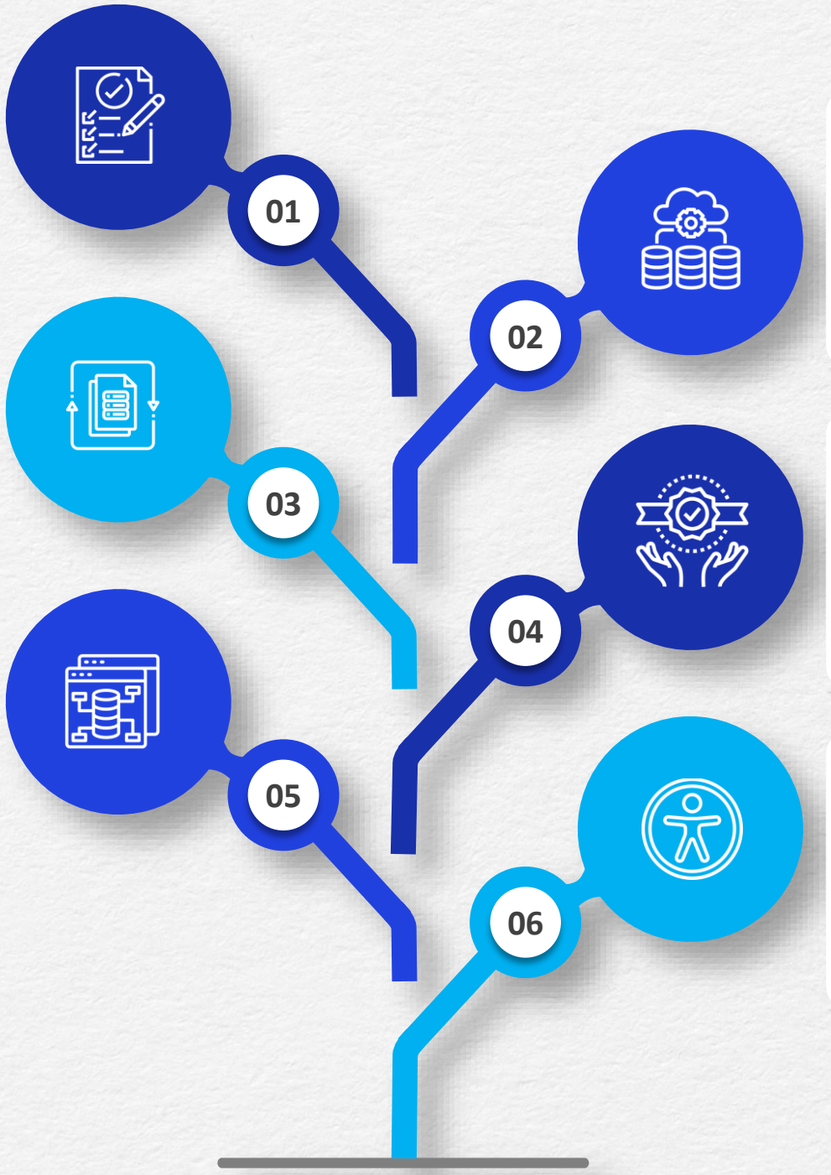
- Low latency / real-time requirement
- Large volumes of data processing

Handling All Types of Data

- Storing & indexing unstructured data is hectic
- Unstructured data makes database operations difficult

Database Management

- Scalability
- Performance maintenance
- Growing complexity in landscape



Managing Data from Multiple Sources

- Different sources, structures and identifiers
- Handling similar data sources and metadata fields

Automated Cleaning & Standardizing

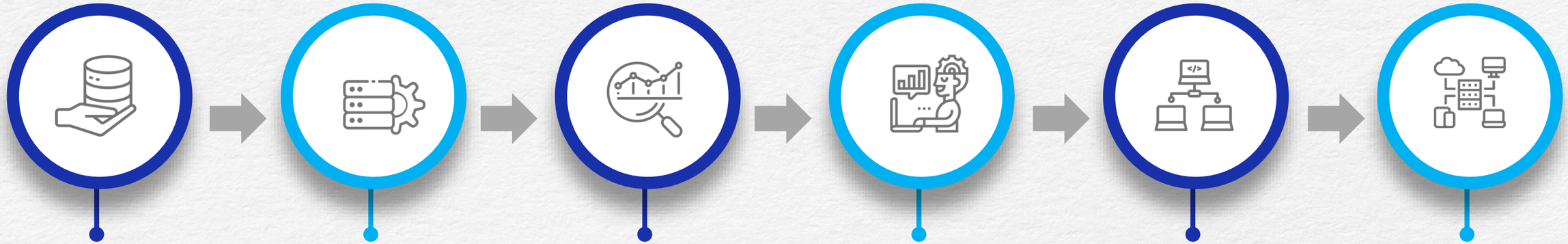
- 70% time of data scientists is spent on data prep
- Lack of blending of the ETL and analytics process

Real-time Accessibility with Intelligent Output

- Eliminating sluggish processes
- Compliance challenges with sensitive data access

The Data Analytics Value Chain

Our Presence Spans Across the Typical Data Analytics Value Chain



Capture

- Data Acquisition / Ingestion
- Various Sources
- Structured / Unstructured
- Numeric / Text / Multimedia
- Setup Once / Automate

Pre-Processing

- Cleaning
- Validations
- Completeness
- Augmentation
- Business Rules
- Data Model / Warehouse

Analyze

- Data Exploration
- Business Logic
- Statistical Analysis
- Feature Engineering

AI-ML

- Data Preparation
- Model Selection
- Model Training
- Model Validation and Testing

Deployment

- Model Deployment
- Ongoing Recalibration

Communicate

- Data Reporting
- BI Data Visualization
- Reports Distribution
- Real-time Reporting

Capturing and Processing Data to Render it Useful

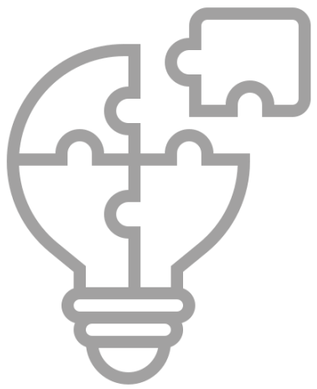
Few Common Challenges

Data Capturing

- Identifying useful data amid the sea of available data
- Integrating data from the abundance of data sources
- Capturing of unstructured data
- Integrating data in different formats

Pre-Processing

- Managing large amount of available data
- Incompatible existing legacy data systems
- Cleaning and validating data is manual and time-consuming
- Achieving data consistency for efficient model building



How Data Science can Help

- Integrate, clean, and orchestrate multiple large data types like strings, integers, symbols, etc.
- Use of AI and NLP to extract data from all document formats

Case Study: Extraction and Ingestion of Large Volumes of Unstructured Data for a SEC Registered Institutional Asset Manager

'Man + Machine' Efforts Resulted in Highly Efficient and Accurate Proprietary Database Creation

Business Requirements

- **Create an Excel-based Database of Tenant Rent Invoices**
- **Converting More than 50K Historical PDFs into Standardized Excel Files**
- **Updating of New Invoices through the Platform**

Challenges

- **Invoices were in Different Layouts and had Non-Standard Terminologies**
- **In Many Cases, the Invoices Had Up-to 400 Pages**
- **PDFs were of Different Types - Text, Scanned Images, and Encrypted, some of which are Extremely Difficult to Process**

Solution Approach

- **DPA Configured a Fully Automated 4-Step Process:**
  - Data Extraction
  - Data Processing
  - Quality Check
  - Final Check
- **Successfully Delivered High-quality Standardized Excel-Based Output Using Multiple Proprietary ML Models**

Business Value

- **600,000+ Tenant Rent Invoice Pages Processed**
- **Scanning Time for a 10 Page File Reduced from 8 hours to less than 2 hours via Automation**
- **Process of Converting 66K+ Historical PDFs, Involving ~640,000 Pages, Completed within 24 months**

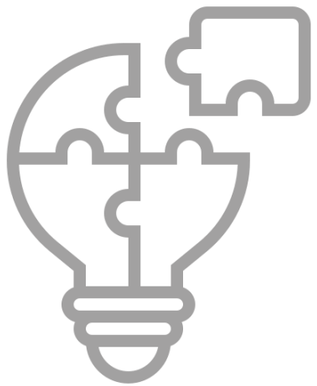
Few Common Challenges

Analyzing

- Time-consuming due to the sheer volume of data
- Identifying and handling anomalies in data exploration phase
- Applying correct business logic to generate insights
- Extracting relevant insights from the vast amount of data, especially unstructured data

AI-ML Capabilities

- Substantial amount of time spent in data preparation
- Handling several different parameters in model building
- Lack of expertise to use relevant libraries properly
- Creating sufficient test cases to validate the models



How Data Science can Help

- Data Science assisted model solutions for significant improvement in data analyst productivity
- Automated solutions to analyze daily fund flows data from multiple sources, capture and cross-reference data using automation tools
- ML-based tracking and extracting of information to shortlist the most relevant articles from prescribed sources with automated PDF generation

Case Study: ML Model to Analyze Offerings from Multiple US Banks

Competitive Intelligence Using AI-ML for a Leading Provider of Technology and Insights to Financial Institutions

Business Requirements

- Track Advertisements and Offers by Leading Banks in the US
- Compare Multiple Data Fields in these Offers
- Highlight the Trends in the Banking Offers

Challenges

- Data Classification
- Unable to Draw Insights from Underlying Unstructured Data
- Identifying and Handling Anomalies in the Data Exploration Phase

Solution Approach

- A Machine Learning Model to Analyze Unstructured Textual Data
- ML-driven Trend and Pattern Analysis of Bank Offers
- Predictions About Possible Offers in Near Future Based on Current Intelligence

Business Value

- Develop New Intelligent Insight Offerings
- Deep Understanding of the Market by Study of Unstructured Data
- Predictive Intelligence to Facilitate Formulation of Business Strategy
- Eliminating the Scope of Human Errors to Influence Output

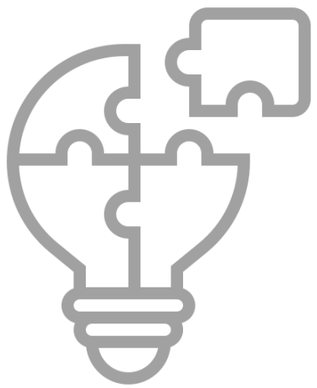
Few Common Challenges

Deployment

- Business Requirements for inference clashing with what data scientists desire for training
- Identifying and resolving model issues post deployment
- Setting up scalable architecture for inference
- Retraining models to account for shifts in a business environment and data patterns

Communicate

- Selecting the right BI and reporting tools
- Delivering real-time, high-quality insights
- Adhering to changing stakeholder reporting demands



How Data Science can Help

- Cloud deployment and continuous monitoring of model
- Systems for automated retraining of models
- Monitoring dashboards offering a comprehensive view of all the data
- Developing an enterprise-wide end-to-end BI and analytics platform

Case Study: Full-Fledged Data Analytics Solution for a Leading Indian Bank

300+ Ad-Hoc Analytics Delivered for key Decision-Makers

Business Requirements

- Design an end-to-end Data Analytics Solution for the Bank with Fast Growth Ambitions
- Deployed the Same with Extreme Fiscal and Temporal Budget Constraints

Challenges

- Handling Complex Data Security and Typical Privacy Issues of Banks
- Deploying Advanced AI and ML Techniques on Limited Hardware Budget Given by the Bank

Solution Approach

- International Best Practices Mapping
- Insights into Technology Landscape Evolution
- Domestic Competitor Mapping
- Fusing all the above into a Coherent Solution

Business Value

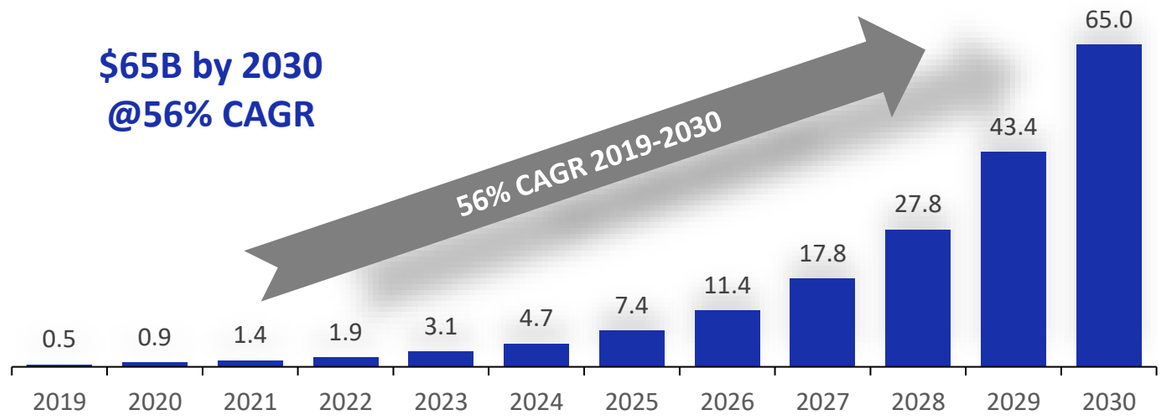
- Detailed View of all Banking Operations
- Internal Data Moved from Silos into a Single Comprehensive View
- Internal Data got Married to External Data to give Competitive Comparison

What we May See in the Not-so-Distant Future

Quantum Computing - Expected to Transform Financial Services

- Quantum Computing (QC) is expected to revolutionize the solving of computationally challenging problems, especially problems of optimization
- The computational power of Quantum Computing has immense possibilities as multiple scenarios can be run simultaneously to arrive at results faster
- Applications of Quantum Computing in the Financial Domain could be:

Global Quantum Computing Market (USD billion)*



Source: Quantum Computing Inc

Portfolio Management

- Simulation of investment options in choosing the optimum Risk return option
- Optimized diversification and rebalancing strategies

Risk Analytics

- Analysis of a greater range of options in a shorter period of time, enabling FIs to better understand risk and prepare for unforeseen events
- Monte Carlo methods are expected to be much faster with QCs

Fraud Detection

- QC's data modeling capabilities are expected to prove superior in finding patterns, performing classifications, and making predictions that are not possible today because of the challenges of complex data structures

Automated Data Handling Systems - A Must Have for Data Scientists

Data Preparation, Cleaning and Validation Takes-up Majority of Data Scientists' Time, Leaving Very Little Time for Model Building and Testing

Current Data Management Available tools in Market are at Best Semi-automated with Very Little Data Cleaning and Validation Features

An Automated System for Handling data Preparation and Related Tasks which Requires Limited Intervention from Data Scientists is the Need of the Hour

SMART PIPE

A Modern ML-driven Cloud-based Platform Providing Access to Comprehensive Data Management Services with Automated Data Ingestion, Cleaning, Advanced Validations and more

SmartPIPE - How is it Different?

01 **100% Automated with Faster One-Time Set-Up**

02 **Easy to Use Interface to Manage all Data in One Place**

03 **Automated Data Type Detection, Data Mapping, and Data Filtering**

04 **Data Formatting, Standardization, Outlier Detection, Deduplication, and Correction**

05 **Designed for Self Service, does not Require Technical Knowledge**

06 **Encrypted User Data and Credentials to Ensure Privacy**

07 **Stores Minimal User Data and Offers Role Based Access Controls**


Q
A

Contact Us

India: Mumbai
Head Office

Click Here

5A- B Wing, Trade Star, Andheri (East), Mumbai - 400 059,
Maharashtra, India
+91 9967066333
info@decimalpointanalytics.com

India: Nashik
Corporate Office

Click Here

Tilakwadi, Nashik - 422 002,
Maharashtra, India
+91 88056 99353

India: GIFT City
Corporate Office

Click Here

D-601, World Trade Center, GIFT City,
Gujarat, 382355, India
+ 91 99670 66333

United Kingdom
London

Click Here

160 London Road, Barking, Essex
IG11 8BB, United Kingdom
+44 20 3239 7604

USA
New York

Click Here

17 State Street, Suite 4000,
New York, NY 10004, U.S.A
+1 917 310 1906

THANK YOU!

www.decimalpointanalytics.com