International Journal of Research in Advanced Engineering and Technology
ISSN: 2455-0876
Impact Factor: RJIF 5.44
www.newengineeringjournal.in
Volume 3; Issue 3; September 2017; Page No. 25-30

Digital disruption in banking: A data science approach for customer experience & innovation in banking

[1] Vamsi Krishna Peddi, [2] Ramesh Mummineni
[1] Chirec International School, Hyderabad, Telangana, India
[2] Data Scientist, Development Bank of Singapore, DBS, Singapore, Malaysia

Abstract
The ability to leverage and profit from new technologies is at the core of future growth and survival in the transformation journey of banking. The digitization of banking is happening at a faster pace than most organizations can handle. Customer expectations are increasing as well. Innovation is playing a critical role.
Data and advanced analytics have arrived. The volume of available data is growing exponentially, with more added every day from billions of phones, sensors, payment systems, and cameras. Machine learning is becoming ubiquitous, but banks are struggling to turn data into value. The big data revolution happening in and around 21st century has found a resonance with banking firms, considering the valuable data, they've been storing since many decades. This data has now unlocked secrets of money movements, helped prevent major disasters and thefts and understand consumer behaviour. Banks reap the most benefits from big data and machine learning as they now can extract good information quickly and easily from their data and convert it into meaningful benefits for themselves and their customers.
Banks internationally are beginning to harness the power of data in order to derive utility across various spheres of their functioning, ranging from sentiment analysis, product cross selling, regulatory compliances management, reputational risk management, financial crime management and much more.
In this article, we are going to identify the current challenges and the need of adoption of new technology by replacing the existing one due to the digital disruption has become essential. It has become imperative for banks to use the Machine Learning and Data Sciences lifecycle for 360 degrees' view of customer experience and for improved decision making.

Keywords: banking, technology, machine learning and data science

1. Introduction
In the second it takes to say "data," people around the world generate about 10,000 tweets, make 1,805 Skype calls, upload five hours of YouTube video and send more than 2.4 million emails. Globally, we produce 2.5 exabytes (or 2.5 billion gigabytes) of data in a day, and IDC research predicts we'll generate 40 zettabytes - that's 40 trillion gigabytes - of data by 2020.
This useful data doesn't generally come neatly packaged in databases. Among this trove of data is knowledge to help banks analyse and fine-tune business processes, create targeted marketing campaigns, efficiently compile reports, comply with regulatory requirements and, in turn, remain competitive and profitable. But how to get at it?
In the last few years, a range of powerful market forces and competitive imperatives have driven banks across the industry to embrace advanced analytics. There is broad consensus that carriers need to adopt and expand analytics programs to improve insight into the business and decision making in everything from product design, marketing and distribution to underwriting and claims. As banks address key strategic, operational and technical issues, they will need to embrace a holistic approach to developing their analytic capabilities. This will require fundamental and coordinated change across every phase of the banking life cycle.
Banking Analytics, or applications of Data Mining in banking, can help improve how banks segment, target, acquire, and retain customers. Additionally, improvements to risk management, customer understanding, risk and fraud enable banks to maintain and grow a more profitable customer base. A number of financial institutions have been quick to recognize and adopt this emerging technology – and it is changing the banking landscape and giving banks and financial institutions previously untapped savings, margins and profit.
The major areas in which advanced analytics and data science practices can be applied in the banking sector are:

1. Spending Pattern of Customers
2. Channel Usages
3. Customer Segmentation and Profiling
4. Product Cross Selling based on the profiling to increase hit rate
5. Sentiment and Feedback Analysis
6. Security and Fraud Management
7. Customer Experience

1.1 Challenges of Current System
Recent research shows that 'organizational silos' are the biggest barrier to success in big data. Dearth of analytics talent, high cost of data management, and a lack of strategic focus on big data are also major stumbling blocks. Finally, privacy concerns – which are high on many bank executives'

agendas – are also a significant issue.

2. Banks are struggling to Profit from Increasing Volumes of Data

More than 70% of banking executives worldwide say customer centricity is important to them. However, achieving greater customer centricity requires a deeper understanding of customer needs. Research indicates that only 37% of customers believe that banks understand their needs and preferences adequately.

This may be surprising given the increasing volume and variety of data that banks have about their customers. The frequent use of web and mobile channels has led to a steady increase in the number of customer interactions and, as a result, increasing volume of customer data. However, banks are only using a small portion of this data to generate insights that enhance the customer experience. For instance, research indicates that less than half of banks analyse customers' external data, such as social media activities and online behaviour. Further, only 29% analyse customers' share of wallets, one of the key measures of a bank's relationship with its customers.

2.1 Silos of Data Block a Single Customer View

Customer data typically resides in silos across lines of business or is distributed across systems focused on specific functions such as CRM, portfolio management, loan servicing, etc. As such, banks lack a seamless 360-degree view of the customer. Further, many banks have in inflexible legacy systems that impede data integration and prevent them from generating a single view of the customer. For instance, Deutsche Bank embarked on a big data project to analyse a large amount of unstructured data, but faced difficulties in the extraction of data from legacy systems, and their integration with big data systems.

2.2 The Skills and Development Gap Needs Closing

Banks need new skill sets to benefit from big data analytics. New data management skills, including programming, mathematical, and statistical skills go beyond what is required for traditional analytics applications. For instance, 'data scientists' need to be not only well versed in understanding analytics and IT, they should also have the ability to communicate effectively with decision makers. However, this combination of skills is in short supply. Three-quarters of banks worldwide do not have the right resources to gain value from big data. Banks also face the challenge of training end-users of big data, who may not be data experts themselves but need to use data to enhance decision-making.

2.3 Lack of Strategic Focus: Big Data Viewed as Just another 'IT Project'

Big data requires new technologies and processes to store, organize, and retrieve large volumes of structured and unstructured data. Traditional data management approaches followed by banks do not meet big data requirements. For instance, traditional approaches hinge on a relational data model where relationships are created inside the system and then analysed. However, with big data, it is difficult to establish formal relationships with the variety of unstructured data that comes through. Similarly, most traditional data management projects view data from a static and/or historic perspective. However, big data analytics is largely aimed to be used in a near real-time basis. While most IT projects are driven by the twin facets of stability and scale, big data demands discovery, ability to mine existing and new data, and agility. Consequently, by taking a traditional IT- based approach, organizations limit the potential of big data. In fact, an average company sees a return of just 55 cents on every dollar that it spends on big data.

2.4 Privacy Concerns Limit the Adoption of Customer Data Analytics

The use of customer data invariably raises privacy issues. By uncovering hidden connections between seemingly unrelated pieces of data, big data analytics could potentially reveal sensitive personal information. Research indicates that 62% of bankers are cautious in their use of big data due to privacy issues. Further, outsourcing of data analytics activities or distribution of customer data across departments for the generation of richer insights also amplifies security risks. For instance, a recent security breach at a leading UK-based bank exposed databases of thousands of customer files. Although this bank launched an urgent investigation, files containing highly sensitive information - such as customers' earnings, savings, mortgages, and insurance policies - ended up in the wrong hands. Such incidents reinforce concerns about data privacy and discourage customers from sharing personal information in exchange for customized offers.

3. Ways to Utilize Advanced Analytics and Machine Learning techniques

3.1 Risk Management

The banking industry is built on risk, so every loan and investment needs to be evaluated. Analytical tools can give banks new insights into their systems, transactions, customers and environments to help them avoid certain risks. For instance, a bank could analyse regional weather data and match that with the age and integrity of the area's buildings to help determine whether to offer insurance in that market. Regional economic data and historic sales data can help determine when a housing market is poised to rebound and where to offer low-interest loans or invest in rebuilding. Banks can analyse the factors that cause borrowers to default on loans and craft new programs to circumvent those factors. It can also make systems more transparent so that institutions can detect internal or external fraudulent activity and identify past patterns to prevent future fraud.

3.2 Marketing and Sales Automation

With the volumes of data available today, banks can gather previously unimaginable information about each of their customers. This gives them a better understanding of customers' needs and helps them to address these needs proactively. It also allows different departments within a bank, such as marketing, sales and IT, to work more cohesively as a single unit. For example, rather than pushing out products to all customers, banks can now merge analytical and sales force automation tools to market products tailored to customers' current situations, whether they're building new homes,

opening small businesses or starting families. In fact, financial services firms using analytical tools reported a 7 percent improvement in cross-sell and up-sell revenues.

3.3 Customer Profitability

Analytics also gives banks up-to-date information on their most profitable customers and the banking choices they make. Banks can use that information to retain high-value customers, market the right products to them and decide which products to invest in for the greatest return.

3.4 Performance Analytics, Budgeting and Product Innovation

Banks can use these tools to measure business and employee performance and then create branch budgets and employee goals based on past achievements. In addition, they can target training and education of these employees for off-peak times and monitor progress toward goals in real time. Banks can also use performance data about products, features and services to create new offerings designed around current customer demand.

3.5 Historical Analysis

Looking at past internal and external data, banks can plan for the future. Data Analytics can help them spot patterns, address issues going forward and set goals that improve upon historic metrics.

3.6 Executive Dashboards

Using graphs, charts and animation, customizable interfaces help users visualize data. Managers can run queries and pull reports based on their needs. They can analyse the percentage of loans by type, monthly operating expenses or profit and loss by region.

3.7 Regulatory Compliance

When asked to name their greatest business concern, 68 percent of bank executives cited regulatory compliance. Data Analytics can help banks gather, organize and analyse data, compile reports and comply with requirements.

4. Current Technologies

4.1 R: The Most Popular Language for Data Science

Once the data scientist has completed the time-consuming process of "cleaning" and preparing the data for analysis, R is a popular software package for actually doing the math and visualizing the results. An open-source statistical modelling language, R has traditionally been popular in the academic community, which means that lots of data scientists will be familiar with it. R has literally thousands of extension packages that allow statisticians to undertake specialized tasks, including text analysis, speech analysis, and tools for genomic sciences.

- Parallel helps R take advantage of parallel processing for both multicore Windows machines and clusters of POSIX (OS X, Linux, UNIX) machines.
- Snow helps divvy up R calculations on a cluster of computers, which is useful for computationally intensive processes like simulations or AI learning processes.
- Rhadoop and Rhipe allow programmers to interface with Hadoop from R, which is particularly important for the "Map Reduce" function of dividing the computing problem among separate clusters and then re-combining or "reducing" all of the varying results into a single answer.

4.2 Java & the Java Virtual Machine

Organizations that seek to write custom analytics tools from scratch increasingly use the venerable language Java, as well as other languages that run on the Java Virtual Machine (JVM). Java has rich open-source libraries for data mining, including Mahout and Weka, and the JVM provides robust memory management and exception handling. Other programming languages that can be used with the JVM include:

- Scala: It is becoming increasingly popular in data mining because it permits developers to use object-oriented programming (OOP) as well as functional programming. Users of Scala include The Guardian, LinkedIn, Foursquare, Novell, Siemens, Twitter, and the SPARK data mining environment at the UC Berkeley AMP Lab.
- Clojure: A dialect of the 1980s-era artificial intelligence language LISP, Clojure is a primarily (although not 100%) functional language that also runs on the JVM. Clojure keeps data static and was designed for running concurrent processes. Users of Clojure include Netflix, Zendesk, Citibank, WalMart Labs, and Spotify.

4.3 Python: A High-Level Programming Language with Excellent Data Libraries

Python is a high-level language, meaning that the creators automated certain housekeeping processes in order to make code easier to write. Python has robust libraries that support statistical modelling (Scipy and Numpy), data mining (Orange and Pattern), and visualization (Matplotlib). Scikit-learn, a library of machine learning techniques very useful to data scientists, has attracted developers from Spotify, OK Cupid, and Evernote, but can be challenging to master.

4.4 Excel: Powerful Data Analytics on a Smaller Scale

Excel can actually accomplish a lot of sophisticated analysis-plus, it's easy to use and widely available. While it's not best for analysing truly massive, unstructured datasets-for example, a massive dataset of some 30 million healthcare records distributed via Hadoop across dozens of servers-it is surprisingly powerful when used for a variety of data analytics projects at a small scale. These can include clustering, optimization, and predictive modelling using supervised AI learning or forecasting techniques.

4.5 SAS (Statistical Analysis System): Data Mining Software Suite

Used for advanced analytics, data management, and social media analytics, SAS is a robust suite that's popular for business intelligence analysis of large data and unstructured datasets. In 2015, SAS topped the Gartner Magic Quadrant list in terms of "ability to execute" in the category of advanced analytics platforms due to the breadth and quality of its

predictive modelling and data mining techniques. With a well-regarded visualization tool and integration with open-source tools like R, Hadoop and Python, SAS also puts significant effort into making tools backwards compatible, an important feature when looking at older historical datasets.

4.6 IBM: SPSS Modeler and SPSS Analytics

Forrester Research Wave ranks IBM's advanced data analytics platform as the top offering in the advanced analytics category for its breadth of tools that handle all elements of big data modelling: loading, "cleaning," preparing, and then predictive modelling, whether using statistical or machine learning techniques. These tools integrate Hadoop to facilitate file-system computing using big datasets. The Social Media Analytics product helps data scientists harvest data from Twitter, Facebook, and other platforms to perform customer sentiment analysis.

Other makers of highly rated commercial tools for advanced data analytics include SAP, Oracle, and Alteryx.

4.7 MySQL: Open-Source RDBMS

MySQL is a widely-used RDBMS (relational database management system) and one part of the LAMP software stack. This free, open-source database management system is used by web applications like WordPress, Drupal, Facebook, Twitter, and YouTube.

4.8 MongoDB

The most popular NoSQL database system available on the market is the open-source MongoDB, which has been used by Met life, The Weather Channel, Bosch, and Expedia. One of the fastest-growing big data projects involving MongoDB is Apache Spark, a distributed computing framework from the Apache Software Project that's designed to operationalize real-time analytics. Paired up with MongoDB, Spark allows organizations to put real-time analytics reporting to use.

Other commonly used open-source NoSQL databases include HBase, Maria DB, Cassandra, SAP, IBM, Enterprise DB, Inter Systems, and Mark Logic.

4.9 Hadoop: File System Computing

Hadoop is an open-source software tool specially designed to help data scientists manage the unwieldiness of big data. It eliminates the need to extract data from the storage devices altogether, bringing the analytics to the data so it can be processed in place. It has increasingly become the industry standard for file system computing projects involving big data, with prominent users including Facebook, Yahoo, and The New York Times.

There are many other platforms that do file system computing, such as Sci. DB, but Hadoop has risen to the top with user contributions that extend its functionality, like Hive, Pig, Spark, and Map Reduce. Even software giants like Microsoft and IBM have created their own Hadoop tools, rather than reinventing the wheel.

4.10 Adobe Target (formerly Test & Target)

With Target, web developers can engage in more elaborate segmentation and custom JavaScript-based experimentation. For example, Target has a particularly nimble interface that makes the cumbersome process of multivariate testing simple and assigns different statistical weights. Clients include Marriott, AOL, and Redbox.

Other testing platforms include: Monetate, A/B Tasty, Qubit, Visual Website Optimizer, and Unbounce.

5. Current Tools

There are a variety of tools available for both enterprise and personal use. The ones listed below are open source and most widely used.

5.1 Tableau Public

Tableau democratizes visualization in an elegantly simple and intuitive tool. It is exceptionally powerful in business because it communicates insights through data visualization. Although great alternatives exist, Tableau Public's million row limit provides a great playground for personal use. In the analytics process, Tableau's visuals allow users to quickly investigate a hypothesis and just go explore the data before embarking on a statistical journey.

5.2 Open Refine

Formerly Google Refine, Open Refine is a data cleaning software that allows developers to get everything ready for analysis. Open Refine contains a number of clustering algorithms (groups together similar entries) and makes quick work of an otherwise messy problem.

5.3 KNIME

KNIME allows analysts to manipulate, analyse, and modelling data in an incredibly intuitive way through visual programming. Essentially, rather than writing blocks of code, users can drop nodes onto a canvas and drag connection points between activities. More importantly, KNIME can be extended to run R, python, text mining, etc., which gives us the option to dabble in the more advanced code driven analysis.

5.4 Rapid Miner

Much like KNIME, Rapid Miner operates through visual programming and is capable of manipulating, analysing and modelling data. It is used for business and commercial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the machine learning process including data preparation, results visualization, model validation and optimization.

5.5 Google Fusion Tables

Meet Google Spreadsheets cooler, larger, and much nerdier cousin. Google Fusion tables is an incredible tool for data analysis, large data-set visualization, and mapping. Not surprisingly, Google's incredible mapping software plays a big role in pushing this tool onto the list. It has the capability to recognise the type of data present and map them accordingly. E.g. Google Fusion tables can recognize the latitude and longitude without explicit specification.

5.6 Node XL

Node XL is a visualization and analysis software of networks

and relationships. Think of the giant friendship maps we see that represent LinkedIn or Facebook connections. Node XL takes that a step further by providing exact calculations.

5.7 Import.io

Web scraping and pulling information off of websites used to be something reserved for the nerds. Now with Import.io, everyone can harvest data from websites and forums. Simply highlight what wanted data and in a matter of minutes Import.io walks through the statistics and "learns" what we are looking for. From there, Import.io will dig, scrape, and pull data for the executives to analyse or export.

5.8 Google Search Operators

Google is an undeniably powerful resource and search operators just take it a step up. Operators essentially allow us to quickly filter Google results to get to the most useful and relevant information. For instance, say we are looking for a Data science report published this year from ABC Consulting. If we presume that the report will be in PDF we can search
"Date Science Report" site: ABCConsulting.com File type: PDF.
Then underneath the search bar, use the "Search Tools" to limit the results to the past year. The operators can be even more useful for discovering new information or market research.

5.9 Solver

Solver is an optimization and linear programming tool in excel that allows users to set constraints. Although advanced optimization may be better suited for another program (such as R's optim package), Solver will make quick work of a wide range of problems.

5.10 Wolfram Alpha

Wolfram Alpha's search engine is one of the web's hidden gems and helps to power Apple's Siri. Wolfram Alpha is the nerdy Google, provides detailed responses to technical searches and makes quick work of calculus homework. For business users, it presents information charts and graphs, and is excellent for high level pricing history, commodity information, and topic overviews.

6. Future Financial Innovations

6.1 Bitcoin Block Chain

The true innovation is not about the currency but is about the bitcoin block chain, a single database massively distributed and replicated, without any central coordinating point. Ripple, a company based in San Francisco, uses this idea to propose what they describe as a low cost, worldwide, multi-currency payment mechanism they say is faster, more secure and simpler to use than any banking based alternative. The system is based on secure wallets managed by the users themselves, and a distributed ledger inspired by the bitcoin block chain. From the customer perspective, the operative word is low cost - the system is open, and as there is no central authority necessary, there are therefore no fees associated to intermediaries.

6.2 Business API

The API has been used for many years by software developers to assemble program components within an application. The new use of APIs is to make business functions available as components on the Internet. The promise of APIs is to grow volumes from existing customers, and attract new customers without friction. As such, making the move to business APIs is a tricky decision to make for financial institutions.

6.3 KYC

The main goal of this is to provide a crowd-sourced, social media based, identity system. In the current financial system, regulation impose that banks and other players conduct “Know Your Customer” (KYC) activities, with the goal of confirming the identity of people that open accounts and conduct transactions. In many occasions though, KYC proves to be very cumbersome because it implies manual checks of various documents. Viewing from the perspective of on boarding billions of people, it is clear that a different solution is needed. OIX (the open identity exchange) and the Respect Network have come up with a crowd-sourced reputation system on the internet, where people on the internet can trust each other based on their digital reputation. For example, Fidor, a financial service provider in Germany, allows an individual to open an account with only their Facebook profile. Of course, they will only let them start with a reduced service, and will provide additional financial services as they know more. This “tiered KYC” system reduces friction and is a powerful instrument of inclusion.

6.4 Open Source Movement

The last innovation is about using the open source movement in the context of financial services. Mozilla is a good example of how it is possible to create a community of millions of developers working on the same project, based on meritocracy. Allevo, a Romanian software company, has recently put in the open-source domain a core banking transaction-processing software.

6.5 A new Generation Technologies

As quickly as past technologies have become the norm, a new wave of emerging technologies will combine digital technologies and the power of data to set new standards. These ‘essential eight’ technologies include:

- The Internet of Things (Io T)
- Artificial Intelligence (AI)
- Robotics
- 3-D Printing
- Augmented Reality
- Virtual Reality
- Drones
- Block chain

Obviously, the prioritization and investment in each of these technologies will vary based on the choice of the banks, business model and strategic goals of each organization.

7. Conclusion

How can we then envision a financial system that rides the wave of change, and leverages the technological innovations mentioned above? Thinking about this from the perspective of the billions of unbanked, and the next billion of hyper-connected individuals that will be the customers of tomorrow. The financial system will need to be a platform that federates existing players (such as banks) and new players (such as telecommunication companies and other information technology companies) that will:

- Be accessible safely and reliably from connected devices, from feature phones to smartphones to other specialized devices.
- Enable new customers to be connected seamlessly, using a crowdsourced and flexible "know your customer" system.
- Enable customers to own their accounts, and operate them themselves or choose freely who to contract to operate it for them.
- Be based on open APIs so new service providers (payments, insurances, loans, etc.) can be connected and discovered easily.
- Be based on open APIs so customers can freely choose how to combine the services on the platform in new ways that suit them.
- In addition, the platform should be available in open source – so that it can be permanently refined and built upon, as a common asset of the humanity.

Another important aspect of the work will be to put all the key stakeholders around the table – banks, central banks, telcos, Start-ups. All hands will be needed on deck to deploy the technologies to work openly, safely and reliably.

In summary, in a world where empowered, always-connected consumers expect to interact with companies in real-time – and no longer rely solely on traditional channels – banks cannot be focused on today and must think about serving the customer of the future, lest they risk irrelevance.

8. References

1. Engler H, Essinger J. The future of banking. UK: Reuters, Pearson Education, 2000.
2. Banks balance technology with brick-andmortar, Rochester Business Journal, https://www.cnbank.com/uploadedFiles/CNB_Site_Home/Your_Bank/News/Articles/2016/RBJ_07%2022%2016.pdf, 2016.
3. Customer Loyalty Statistics: 2015 Edition, Access Blog, Brandon Carter, http://blog.accessdevelopment.com/customer-loyalty-statistics-2015-edition#bank, 2015.
4. Collinson Latitude. Taking Account: The consumer perspective on loyalty programmes in financial services, 2015.
5. Kohavi R, Provost F. Glossary of terms, Machine Learning, 1998; 30(2-3):271-274.
6. David Turner, Micheal Shroeck, Rebecca Shockley. Analytics: The real-world use of big data in financial services. IBM Institute of Business Value in collaboration with Said Business School, University of Oxford; 2013.
7. IOSCO Research Report on Financial Technologies Fintech, https://www.iosco.org/library/pubdocs/pdf/IOSCOPD554.pdf
8. Research Challenges in Financial Data Modeling and Analysis, http://srdas.github.io/Papers/2017_R1_ResearchChallengesBigDataInFinance.pdf
9. Langley Pat. The changing science of machine learning". Machine Learning. 2011; **82**(3):275-279.
10. David JC, MacKay. Information Theory, Inference, and Learning Algorithms Cambridge: Cambridge University Press, 2003