

Big Data at American Express

Asala Aljazaery

7 March 2022

University of Sunderland

Contents

1	Introduction	3
2	Applications of Big Data	5
3	Challenges and Opportunities	10
4	Strategic and Operational Uses of Big Data	12
5	Professional and Ethical Requirements	18
6	Conclusion	20
	References	23

1 Introduction

This report is an overview of American Express and its aims, how it is using big data to achieve its ambitions and a discussion of its critical applications of big data within the organisation. First, the report will discuss what type of data American Express collects and its sources, followed by an understanding of techniques used to enhance the value of big data. Second, challenges and potential opportunities will be presented, followed by a summary of strategic and operational uses of big data and other essential properties. Finally, this report will look at the American Express code of conduct for data protection and privacy principles and other future insights for using big data in American Express.

American Express (New York, United States) is a multinational company that issues its clients with a variety of financial and travel services. It is a market leader for credit cards used for personal, corporate and travel services, and it operates in >40 countries. It is well known for its rich history of utilising data analytics to create and retain a deeper relationship with potential and expected customers. Machine learning has allowed the company to take full advantage of the available datasets within the company (Tikkanen, 2022). The main aim of the business is to be dominant in the payment and customised service industry. American Express believes in investing in new technologies, and it is successful in being a data-driven organisation (Hughes Education, 2022). The company uses its vast data flows to develop applications connecting cardholders with products or services, and it processes billions of transactions (Harvard Business School, 2022). The essential goal of any bank is to protect its customers' finances. Thus, banks are always working to achieve an efficient way of monitoring against any cyberattacks (Galea-Pace, 2022). However, data breaches are a trend impacting risk management, and every second,

72 data records are lost or stolen (Tollefson, 2022). At the same time, data analytics can effectively detect and prevent cybercrime. Therefore, data is money and protecting this data will bring the company value. Furthermore, data risk has to be considered as crucial as financial risk, and companies need to protect information security like they protect their financial assets.

2 Applications of Big Data

The main aim of using big data within American Express is to build a trusted transaction company that makes it easier for their customers to process any payment and protect their customers from any finances fraud (Galea-Pace, 2022). However, this interaction between company and customer has helped the company to make a 360-degree view about the customer. This view was used for making changes in product design and marketing. With the richness of data out there, it is becoming easier for fraudsters to steal peoples' identity credentials and mimic a user. In addition, data analytics was effective in detecting cybercrime and preventing it before it occurred (Ashley, 2020). American Express has ~110 million credit cards in active use, and greater than one trillion dollars of processed transactions to date. These data were fed to the machine learning models to extract the information. American Express handles 25% of US credit card activity, which means a vast amount of data is collected about the cards holders (Galea-Pace, 2022).

American Express has been utilising data to provide their cardholder with applications. These applications connect cardholders with products, or services with benefits and give an incentive to attract more businesses to accept American Express (Galea-Pace, 2022). Currently, American Express is a global leader in payments and personalised services (Hughes Education, 2022).

Big data has a significant weight on the decision-making process within American Express and has been used to detect fraud, bringing merchants and customers closer (Galea-Pace, 2022). American Express uses big data with machine learning models to attract new clients using online marketing and this has led to ~40% increase in the acqui-

sition of new clients as a result of online applications, and ~90% through E-mail (Harvard Business School, 2022). American Express has also been using data to track fraudulent activity, so it can decide to process the transaction or deny it in seconds (Galea-Pace, 2022). Moreover, this process has minimised the company's loss of fraud. American Express has used big data to recommend products and services to their customers based on their lifestyle and buying habits. So it predicts cardholders spending amount and suggests for the cardholder new offers. For example, it can recommend a particular type of restaurant to a cardholder depending on their previous choices of restaurants, and it gives extra advantage by giving these recommended options based on the cardholder geographic information (Harvard Business School, 2022).

A massive number of data points are collected about American Express cardholders, which includes credit card holder and merchant information (Galea-Pace, 2022). In addition, American Express has an enormous volume of structured and unstructured data sets (Ashley, 2022). The company collects data in three ways:

1. Clients give data directly by signing up for products and services.
2. Using the website or the app of American Express, also making any purchase using an American Express card for Electronic payment.
3. Public data that is available from online platforms or databases (American Express Company, 2022c).

American Express uses the closed-loop system in its own network that processes the transactions between buyers and sellers, so it has access to real-time data of both customers and merchants. Real-time data has given the company more advantages than its competitors in the industry. American Express uses recommendation systems with big data to target customers with personalized offers (2015). Therefore, American Express uses predictive modelling. The predictive model analyses any transaction that has been done in the past, which includes ~115 variables to predict any potential customer attrition (Hughes Education, 2022). Fraud detection was performed by using cardholder and merchant information, and spending patterns were connected with deep learning models

in real-time with the ability to detect transactions with a high likelihood of fraud (Harvard Business School, 2022). American Express uses two deep learning-based models depending on the application it is trying to establish, such as:

1. Recurrent neural networks (RNNs) are deep learning neural networks that utilise the NVIDIA TensorRT framework and Triton Inference servers for fraud detection. RNNs are good at predicting future occurrences in datasets. Moreover, this model was used for regression and classification (Ashley, 2020).
2. Generative adversarial networks (GANs) are a form of deep learning neural networks that generate artificial data based on a limited population of data segments. This network is used when new products are launched. GANs produce additional data, and this generated data is used to train and build more accurate deep learning models. Synthetic data is annotated information, which is generated by computer simulations or algorithms and used as an alternative to real-world data (Ashley, 2022).

For using RNNs with big data for fraud detection, the data is preprocessed with these two fundamental steps: (1) feature selection, and (2) feature extraction. These steps are required for fraud detection systems to remove duplicative and nonessential features from the dataset. Moreover, decreasing the resources required for computational power without negatively impacting the model's predictive accuracy.

The main goal of using feature selection on the dataset is to examine the contribution of features in the prediction models and select features that are relevant. Also, a technique called Synthetic Minority Oversampling Technique (SMOTE) was used to optimise fraud classification performance. Credit card fraud detection depends on analysing cardholders' spending behaviour. This spending profile uses a selection of variables that best represents the distinctive behaviour of a credit card, and detects unusual transactions within a customer's purchase. The optimal selection of variables that significantly differentiates both profiles is required to classify credit card transactions efficiently.

For the second step, the method of feature extraction using the principal component analysis (PCA) is used for dimension reduction to maximise variance. Then, uniform

manifold approximation and projection (UMAP), a graph-based algorithm, is used for the emerging dimensionality reduction technique. As a result, The first phase represents high dimensional data points in a weighted neighbour graph. The second phase represents the k-dimensional eigenvectors of the UMAP graph, which is used to represent each point of the original dataset.

Finally, long short term memory (LSTM) is a deep learning recurrent neural network (RNNs) used to model time-series information and evolves so that the transaction label will be predicted based on the previous transactions to finally process the transaction or flag it (Figure 2.1) (Benchaji *et al.*, 2021).

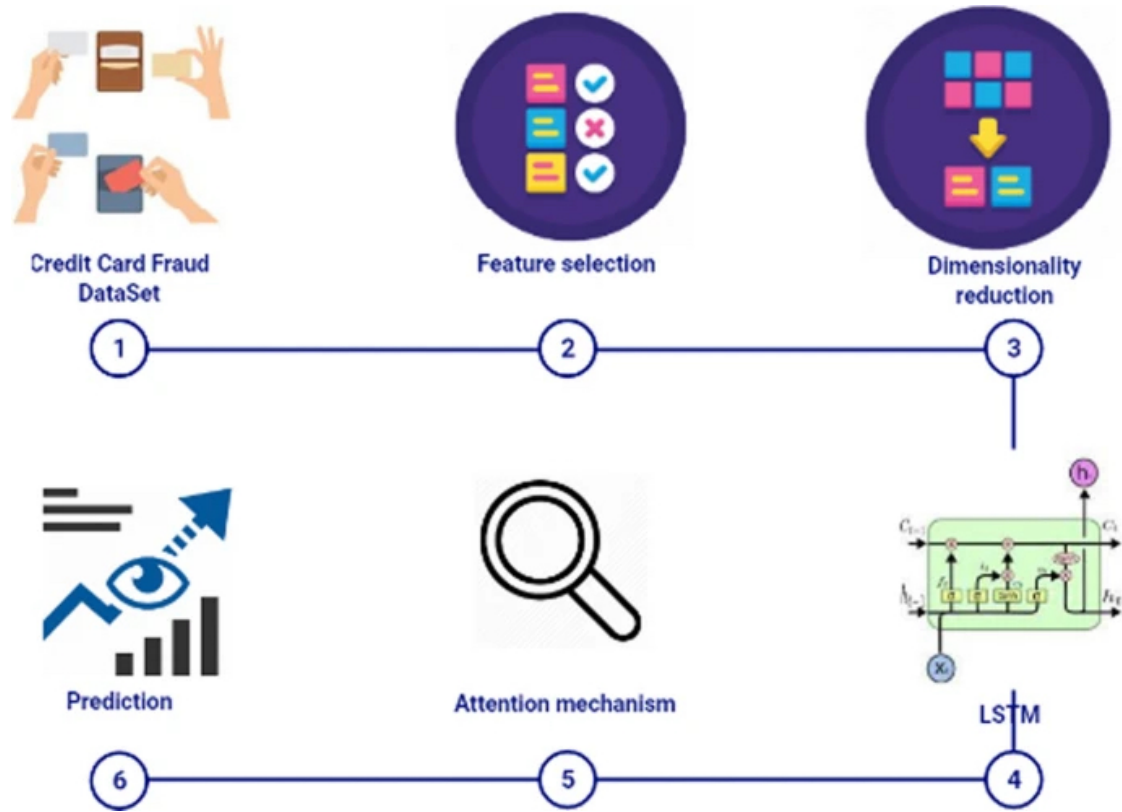


Figure 2.1: Long Short term memory(Benchaji *et al.*, 2021).

Many variables are accumulated from customers and merchants, such as:

1. Customer interaction through mobile-based applications.
2. Payments and e-commerce to analyze shopping trends preferences, preferred mer-

chants and other purchase preferences such as; location, time and date.

3. Non-traditional data sources include phone numbers, cookies and digital data (Hughes Education, 2022).

3 Challenges and Opportunities

The challenges of using big data for data analytics in American Express involve the following:

1. Data capabilities to use on a world wide scale (Bean, 2022).
2. The company was struggling to recruit qualified data scientists (Hughes Education, 2022).
3. Traditional data warehouses in the company were not coping with massive data and machine learning algorithms (2015).
4. Finding the balance between immediate needs and significant data investments (Harvard Business School, 2022).
5. Driving internal efficiencies and cost reductions.
6. Adoption of new technologies requires transformation (Bean, 2022).
7. Shortage of datasets that are suitable for fraud detection models (Ashley, 2020).
8. Data was limited due to the need for customer privacy and trust (Ashley, 2020).

These challenges have impacted the effectiveness of using big data within American Express. However, the shortage of qualified people was the biggest challenge that even other organisations still have today. Talented and trained employees must understand the data and address the root causes of any future risks. Therefore, the team must perform in a cross-functional manner to interpret the risk of customer attrition. American Express team was extremely patient in their experimentation. However, competitors in the market are continuously changing, so the firms need to pivot quickly in response to a very dynamic market (Bean, 2022).

To tackle the transformation challenges in the company, the American Express team started by focusing on the easy-to-solve problems in risk management. It was able to process large volumes of data which has improved the assessment of risk. For example, in a published report involving a machine learning system, thousands of data points were processed in less than two milliseconds and this has saved ~\$1 trillion across the American Express network annually (Bean, 2022). For a better and more effective data processing system, they started using Hadoop and implementing a MapR data platform designed to store, stream and facilitate search on data, though they used a NoSQL database to allow large-scale machine learning applications for further services that drive revenues for the firm (2015). American Express established a big data and cloud computing technology centre in (Silicon Valley, United States) to bring new talents and skills. However, data science and big data remain a significant challenge for any company, so they encouraged decision-makers to use big data by making all the tools and techniques readily available to them. Finally, the shortage of datasets was solved by introducing deep learning neural networks (GANs) and applying some pre-treatment for data to generate datasets from the small datasets (Ashley, 2022).

4 Strategic and Operational Uses of Big Data

American Express employed many machine learning algorithms for various uses, mainly in fraud detection and customer churn prediction (2015). The company has had to change the infrastructure to implement Hadoop as a tool for big data instead of the traditional database warehouses and applied machine learning algorithms, to store and process data for the detection of fraud, new customer acquisition, and recommendations (2015). In addition, American Express has partnered with other businesses such as Uber and Airbnb to accept American Express cards. This has allowed clients to gain rewards every time they use it with these platforms.

Risk2020 is an analysis technique used based on big data that predict how the economy will evolve in the next few years. Depending on that, decision makers improve or change areas in the company to meet the requirement if needed; this will help the company stay in good position within the economy and prepare for any changes up ahead. This approach was met by using other technologies such as cloud, deep learning, mobile computing, and artificial intelligence (Dexlab, 2022). Depending on this prediction, the company will make all future decisions and bring stability to its various transformation.

For merchants, American Express has changed its reports to online trend analysis using anonymous data to help merchants understand how they are doing compared to others (Harvard Business School, 2022). For example, American Express has made a data-driven collaboration with data streams and advertisement companies to make the required actions for online marketing because American Express acquisitions teams are needed to have better insights into the approval process in credit card applications.

These insights were dependent on data points such as the landing pages, submissions of credit card applications and the approval of credit card applications. Therefore, for the collaboration, the companies collected data from two sources:

1. Global marketing campaign performance; because it has a significant level of detail about the online campaigns in American Express.
2. Adobe analytics for real-time analytics and marketing channels.

Then they started to configure new data models used for visualisation purposes by connecting this data models in Tableau for visualisation purposes (Figure 4.1).



Figure 4.1: Visualisation samples in Tableau (Data Streams, 2022).

Developers then kept updating the dashboards for American Express users' needs and gave them insights to optimise their campaigns. American Express acquisitions teams can log at anytime to see their up to date data/reports, which are trusted and easy to understand (Data Streams, 2022).

American express has launched American Express Advance, which is a platform for predictive analytics aimed for corporate clients that provides customised services to help understand their clients' behaviours (Harvard Business School, 2022). Cornerstone is another application that collects online review sites and metrics as data about the per-

formance to track the brand repetition online and build customer engagement to achieve best practices in customer service (Taylor, 2022). All data insights and information provided on all platforms were based on anonymous data for businesses to protect clients' privacy.

In the past, American Express had data warehouses for each service based on its relational database. But now, it uses a single system referred to as the "sync platform". In addition, American Express operates two Hadoop clusters. The first is used for research and development purposes which is distinct from the "sync platform", and the second is used for one-time analytics (Morgan, 2014).

As long as American Express uses predictive analytics, the data availability and the validity of data will be crucial factors. American Express needs to collect the correct quantity of data with good quality of payments data to extract a good value. However, new approaches must be taken to see the bigger picture of client behaviour. For example, American Express may collaborate with other corporates to collect more data about clients or gather information by asking the consumers to answer some questions about the activity they have done (Harvard Business School, 2022).

An important step that determines the success or failure of machine learning techniques is the preprocessing of data. Financial data usually contains features such as different types, ranges, distributions, or missing values. All these data features have proposed challenges within the data processing step, and they should be addressed to ensure adequate training and testing for the model's performance (Efimov *et al.*, 2020). For training GANs models, a study found some preprocessing steps for data that allowed adequate training and testing and enhanced the performance of the data within the model. These steps were:

1. One-hot encoding for categorical features.
2. Missing value indicator feature, which gives one or zeros in dataset where there is a missing value.
3. Box-cox transformation.
4. Standard scaling or min-max scaling.

5. Imputing missing values (Efimov *et al.*, 2020).

After using GANs to generate data, American Express has an internal tool called DataQC, developed to evaluate the data. Hence, it allows users to compare two datasets and gives all the similarities and differences. Furthermore, this tool is used to test data quality and compare features between both datasets and includes: (1) means, (2) rates of missing values, (3) uni- and multivariate distributions and (4) extreme values (Efimov *et al.*, 2020). After that, a t-distributed stochastic neighbour embedding algorithm was used to visualize the overall comparison between datasets. It shows most clusters and gaps accurately in shape and density (Efimov *et al.*, 2020). Another use of big data within American Express is infographics which reflects their data to consumers or decision-makers, so they can further understand the trends, needs or challenges (Amex Trendex, 2022). See Figure 4.2 and Figure 4.3.

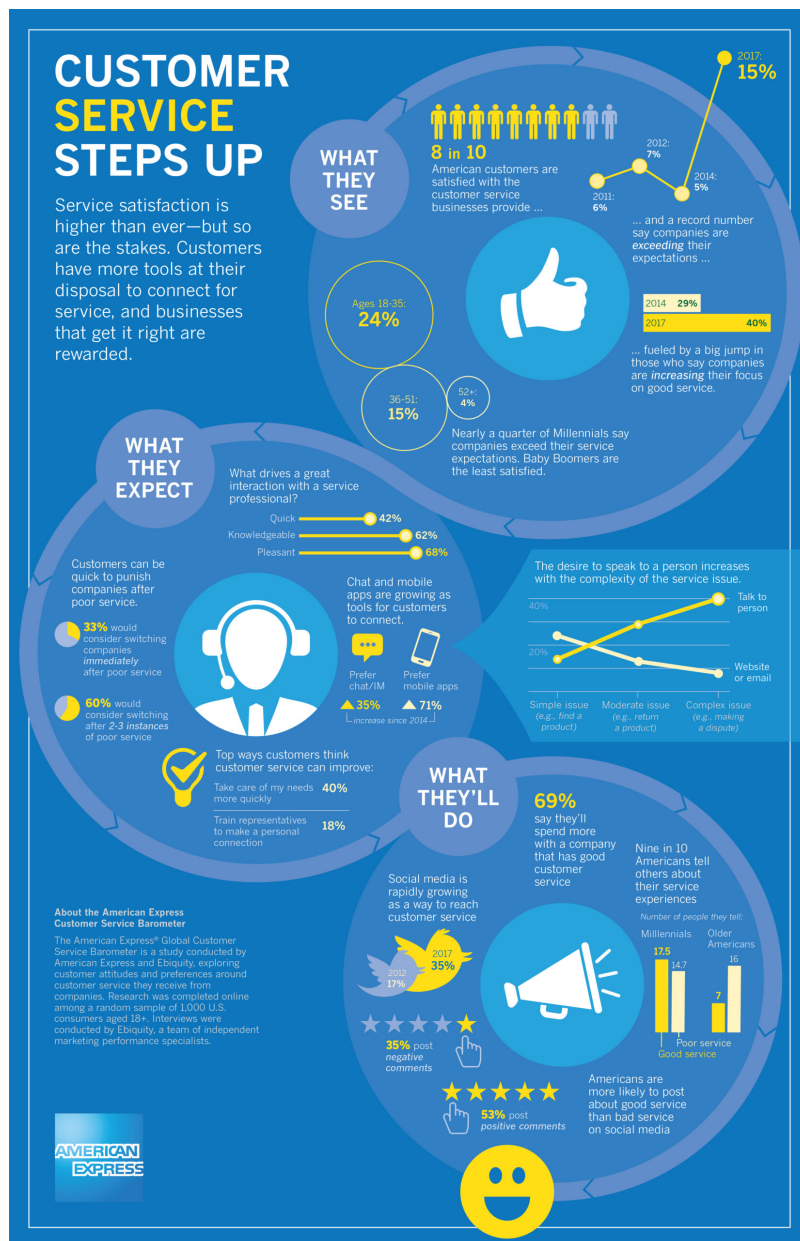


Figure 4.2: American Express customer service infographic scaled (Amex Trendex, 2022)

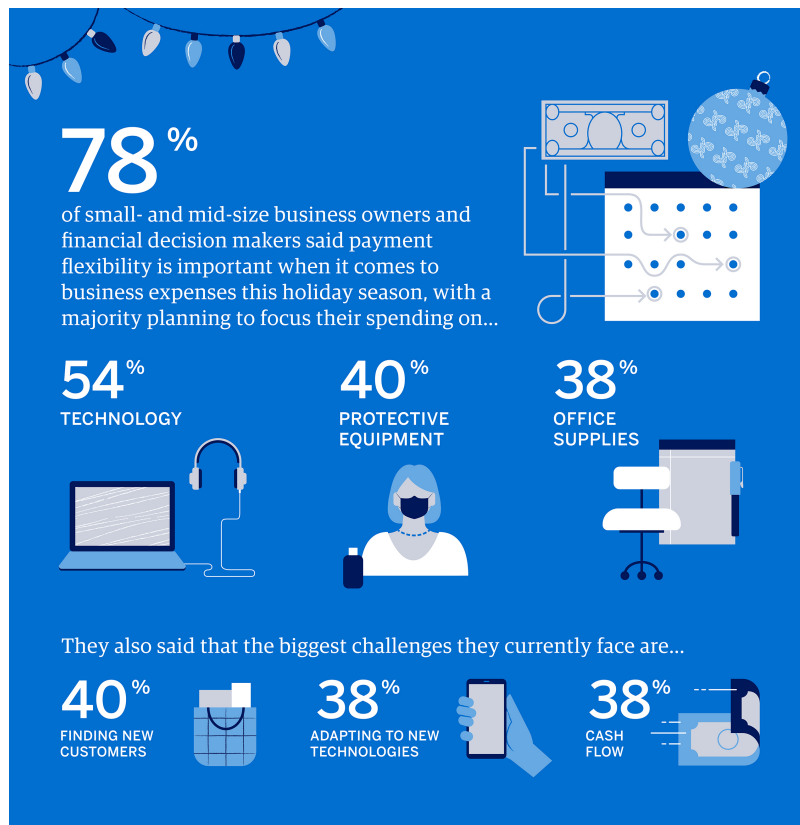


Figure 4.3: American Express Project Charm Infographic (Amex Trendex, 2022) .

5 Professional and Ethical Requirements

In American Express, privacy is paramount, and all the transformations were made under a heavily data regulated industry (Bean, 2022). Employees are responsible for customers' privacy and the confidentiality of customer information; therefore, newly employees at American Express receive training to get a comprehensive understanding of the company and clients' privacy and put it into practice every day to ensure no data breaches happen. American Express has a code of conduct for the employees. These principles ensure that when employees collect customer information, they only collect necessary, accurate information and keep it confidential. This principle includes all types of data, whether this information was collected on paper, electronic or verbal (American Express Company, 2022a).

American Express has data protection and privacy principles based on the following:

1. **Collection:** American Express will collect data in a lawful, fair manner and when needed only.
2. **Notice and processing:** cardholders will be informed about who will process their data and which companies in the American Express group are responsible for that process. Therefore the data will be processed for the purposes that the cardholders have agreed upon.
3. **Choice:** the cardholder can choose if the personal data will be included or removed from marketing purposes lists.
4. **Data Quality:** American Express uses appropriate technologies and well-selected employees. So, individuals' data will be processed promptly and accurately, thus data will not be kept for longer than is necessary.

5. **Security and confidentiality:** Data will be stored confidentially with limited access.
6. **Data sharing:** Data will only be shared with third parties if necessary and authorisation is required by the cardholders.
7. **Openness and Data Access:** The cardholders have the right to ask about how their data will be processed and their rights under these principles.
8. **International Transfers:** Cardholders will be informed if their data is transferred outside of their country. Whatever country the data will be transferred to, these principles will be protected.
9. **Responsibility:** Employees are trained to follow these principles, and they will be reported if a violation occurs (American Express Company, 2022b).

6 Conclusion

The company's nature has an opportunity to access all the clients' data from both sides, customers and merchants' side of the business, resulting in millions of sellers and buyers. American Express uses machine learning across a vast range of data collected about their cardholders to understand customers' behaviour better and to enhance customer service and ensure high security. On the other hand, the company has made a significant effort to adopt new infrastructure and technologies to implement systems that can process large volumes of data. Big data has supported American Express to do more research and analysis, so they can improve clients' experiences where respect for information privacy is paramount. Big data has also improved security by preventing and detecting fraudulent activities (2015). Big data has become centric for American Express and its effective marketing, services and risk management (Hughes Education, 2022). The data scientist teams in the company consist of 800 employees around the world and has achieved the lowest fraud loss rate in the industry (Bean, 2022).

As credit cards have become a common method of payment, credit card fraud has increased. It is therefore crucial that financial institutions continue to improve detection of fraud in an effort to decrease financial loss (Benchaji *et al.*, 2021). Furthermore, to keep the outstanding brand value of American Express, it needs to stay up to date with all digital transformations (Harvard Business School, 2022). The company needs to continuously improve its applications and continue to evolve with the rapidly changing consumer market. This is evident in the fact that it has launched a big data laboratory for research and analytics purposes, which has >25 researchers on the team, >80 patents and >140 papers on Hadoop, machine learning, web databases, and real-time search

(Morgan, 2014).

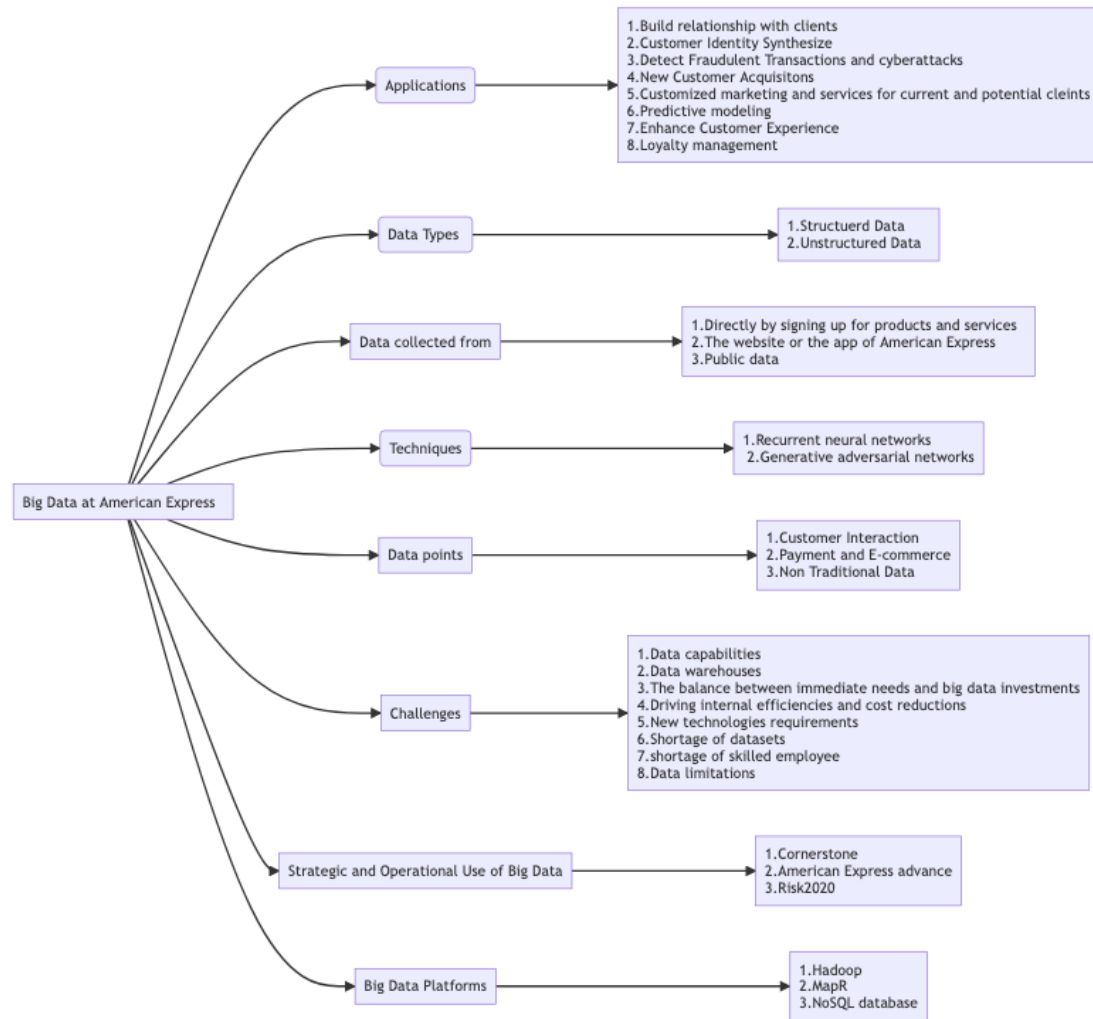


Figure 6.1: Summary

References

- American Express Company (2022a) ‘American Express Company Code of Conduct’. American Express. Available at: http://media.corporate-ir.net/media_files/irol/64/64467/2007_axp_coc.pdf (Accessed: 27 February 2022).
- American Express Company (2022b) *American Express Data Protection and Privacy Principles*. Available at: <https://www.americanexpress.com/icc/data-protection-and-privacy-principles.html> (Accessed: 26 February 2022).
- American Express Company (2022c) *American Express Privacy: Data Collection, Use and Sharing*. Available at: <https://www.americanexpress.com/us/privacy-center/> (Accessed: 5 March 2022).
- Amex Trendex (2022) *Amex Trendex: Small Businesses Find Payment Flexibility to be Critical as They Gear Up for the Holiday Season and 2021*. Available at: <https://about.americanexpress.com/all-news/news-details/2020/Amex-Trendex-Small-Businesses-Find-Payment-Flexibility-to-be-Critical-as-They-Gear-Up-for-the-Holiday-Season-and-2021/default.aspx> (Accessed: 2 March 2022).
- Ashley, J. (2020) *American Express Adopts NVIDIA AI to Help Prevent Fraud and Foil Cybercrime*. NVIDIA Blog. Available at: <https://blogs.nvidia.com/blog/2020/10/05/american-express-nvidia-ai/> (Accessed: 1 March 2022).
- Ashley, J. (2022) *American Express Uses Deep Learning for Better Decision Making*. Available at: <https://blogs.nvidia.com/blog/2019/12/10/american-express-deep-learning/> (Accessed: 5 March 2022).
- Bean, R. (2022) *Inside American Express’ Big Data Journey*. Forbes. Available at: <https://www.forbes.com/sites/ciocentral/2016/04/27/inside-american-express-big->

data-journey/ (Accessed: 1 March 2022).

Benchaji, I. *et al.* (2021) ‘Enhanced Credit Card Fraud Detection Based on Attention Mechanism and LSTM Deep Model’, *Journal of Big Data*, 8(1), p. 151. doi:10.1186/s40537-021-00541-8.

Data Streams (2022) ‘A Data Collaboration Between Datastreams, Our Partner Adversitement and Their Customer American Express’. Available at: <https://www.datastreams.io/wp-content/uploads/2019/05/Client-Success-Story-AMEX.pdf> (Accessed: 2 March 2022).

Dexlab (2022) *How American Express Uses Data Analytics to Promote a Data-Driven Culture*. DexLab Analytics | Big Data Hadoop SAS R Analytics Predictive Modeling & Excel VBA. Available at: <https://www.dexlabanalytics.com/blog/how-american-express-uses-data-analytics-to-promote-a-data-driven-culture> (Accessed: 28 February 2022).

Efimov, D. *et al.* (2020) *Using Generative Adversarial Networks to Synthesize Artificial Financial Datasets*. Available at: <http://arxiv.org/abs/2002.02271> (Accessed: 2 March 2022).

Galea-Pace, S. (2022) *How American Express Uses Big Data to Transform Operations*. SupplyChain. Available at: <https://supplychaindigital.com/technology/how-american-express-uses-big-data-transform-operations> (Accessed: 25 February 2022).

Harvard Business School (2022) *American Express: Machine learning for Customer Churn Prediction and More Effective Customer Retention*. Technology and Operations Management. Available at: <https://digital.hbs.edu/platform-rctom/submission/american-express-machine-learning-for-customer-churn-prediction-and-more-effective-customer-retention/> (Accessed: 28 February 2022).

Hughes Education (2022) *American Express is Redefining Traditional Banking with Data Science*. Available at: <https://www.hugheseducation.com/blogs/american-express-is-redefining-traditional-banking-with-data-science> (Accessed: 28 February 2022).

Morgan, T. (2014) *Why Hadoop Is The New Backbone Of American Express*. EnterpriseAI. Available at: <https://www.enterpriseai.news/2014/10/17/hadoop-new->

backbone-american-express/ (Accessed: 1 March 2022).

RTInsights Team (2015) *How AmEx Found Gold With Machine Learning*. RTInsights. Available at: <https://www.rtinsights.com/american-express-recommendation-engine/> (Accessed: 28 February 2022).

Taylor, I. (2022) *Making Customer Service the Cornerstone for Your Business*. Business Class: Trends and Insights | American Express. Available at: <https://www.americanexpress.com/en-us/business/trends-and-insights/articles/making-customer-service-the-cornerstone-for-your-business/> (Accessed: 6 March 2022).

Tikkanen, A. (2022) *American Express Company History & Facts Britannica*. Available at: <https://www.britannica.com/topic/American-Express-Company> (Accessed: 5 March 2022).

Tollefson, R. (2022) *Risk Management 101: How to Protect Your Company From Fraud*. Business Class: Trends and Insights | American Express. Available at: <https://www.americanexpress.com/en-us/business/trends-and-insights/articles/risk-management-101--how-to-protect-your-company-from-fraud/> (Accessed: 4 March 2022).