

AI STUDIO FOR HEALTHCARE STARTUPS

BigR.io is a US based technology and consulting company with its headquarters in Boston. We empower our clients to drive innovation and advanced analytics through the use of data science. Our expertise in software engineering, product development, database technologies, cloud engineering, web and mobile applications development has helped us to create some highly specialized services and solutions using AI, ML and NLP proprietary tools.

BigR.io stands out as a thought leader in AI advancement with its Machine Learning practice, delivering custom solutions to its clients on the same robust and scalable engineering platform that backs its software and integration offerings. We excel in a complete portfolio of Data Science expertise.

We have now launched an AI Studio specifically for the US based Healthcare startups with AI centricity. We are partnering with startups developing healthcare solutions with AI at the core of their solutions. Our mission is to help you scale and gear up to stay one step ahead from the rest and emerge as winners in the respective domain.



The advantages available to the startups:

- Access to top level talent pool including business executives, developers, data scientists and data engineers.
- Assistance in development and testing of the MVP, Prototypes and POCs.
- Professional services for implementation and support of Pilot projects.
- Sales and Marketing support and potential client introductions.
- Access to private capital sources.

RAPID PATH TO EARLY RESULTS

Startups face numerous challenges when it comes to demonstrating their value proposition, especially when it comes to advanced analytics. Oftentimes, they don't have access to meaningful data, and when they do, they may lack adequate infrastructure to house them. Data quality is always a concern that baffles even well-established companies. While one could argue that social media data is free for all, these unstructured data usually are not presented with desired labels (imagine if you are interested in detecting fake news as a topic). Even structured data collected during normal operations can be questionable in their trustworthiness. Our experience in auto predictive maintenance, for example, discovered that diagnosis of component failures in dealerships can often be partially based on folklores, due to the high billing rate of labor. BigR.io excels in overcoming such initial hurdles which present nearly insurmountable obstacles to a startup operation.

MINIMUM DATA INFRASTRUCTURE

Our data storage solution for small business and startups is based on a combination of a single relational database, possibly in combination with a minimum size Data Lake for accommodating unstructured and high variety data, leveraging the schema-on-read flexibility. Major cloud platforms all provide very low-cost storage offerings. The BigR.io team selects the best fitting platform based on the overall client needs. Some may emphasize NoSql solutions,



others will favor certain managed analytic services. The reliance on Deep Learning vs. AutoML, vs open source libraries all weigh in during our evaluation.

Once raw data from multiple sources are gathered, a company can perform all types of Data Science projects with them.



UNSUPERVISED AND SEMI-SUPERVISED LEARNING

Brute force crowdsourcing approach to data labeling is taxing enough even for big enterprises. For a startup operation, the expenses and efforts are easily prohibitive. While Deep Learning exercises involve sample sizes in the millions, even Machine Learning models will require a minimum of 10's of thousands of samples. Beyond the sheer numbers, the quality of labeling results is far from assured on the Mechanical Turk platform without a well thoughts of cross-checking and validation methodology.

Bigr.io leverages the state-of-the-art semi-supervised techniques to enhance classification accuracy by a factor of 10 to 100 times, thus reducing the amount of labeling to below 1000 samples per experiment. Oftentimes, a PoC (Proof of Concept) project can extract great values from data even without labeled data, by means of unsupervised learning, such as clustering and visualization, an area Bigr.io has a long established credential as an innovator.





Visualization technique applied to unsupervised clustering is used to investigate the effectiveness of battery charging cycle data for classifying battery aging behaviors.

BAYESIAN APPROACHES

The best approach to a business problem often goes beyond the well known Deep Learning and Machine Learning. The Bayesian school of statistical studies is still the most illuminating approach when it comes to getting a complete map of the underlying distribution, from which any insights can be derived, even as an afterthought. Many people in the analytics industry are familiar the basic formula:



An example would be, given a collection of spam and non-spam emails, can you predict the likelihood that the word "congratulations" serves as a trigger word. Advanced Bayesian modeling was first enabled during the Manhattan project with the advent of the ENIAC computer. In today's world of unlimited computing power, techniques such as Gibbs Sampling can be efficiently applied with a small budget for a GPU machine on a cloud platform, or TPU in case of Google Cloud.

The wonders of Bayesian modeling go beyond distribution mapping for observed random variables. As often is the case, the hidden variables are the usually most powerful predictors. Think, for example, how a business would model the readiness of a consumer to pull the trigger on a shopping cart. While no single web analytics parameter such as time spent on a page can directly predict this readiness, a Hidden Markov Model can capture such a tendency as a numerical value, as well as highly valuable insights such as the Next Best Action to maximizes the effectiveness of a marketing campaign.

Bayesian techniques can measure hidden values which make powerful predictors such as what is the Next Best Action to take to most effectively communicate with an individual prospect





SOCIAL LISTENING

One of the best sources to gather market feedback for startups is to simply follow social media, when a dedicated focus group is out of budget reach. While the opinions are freely available, the challenge is to lift the truthful signals

from the noise, including some intentionally injected by trolls. Neural NLP techniques can now be used to perform abstractive summarization, sentiment analysis, entity and relation extraction, and machine translation. Word embedding based topic modeling yields faithful numerical representations of natural language statements that leads to much improved classification accuracy. Our big data team excels in setting up infrastructure for ingesting social media content and preparing them for business insight analysis.

Social listening is the most direct and economical way to get insights.



MASTERING AGILE

BigR.io builds all projects utilizing an "Agile without Religion" philosophy. Our flexible approach is based on the original text and intent of the Agile Manifesto, foregoing heavyweight "agile" methodologies aimed at larger, less flexible organizations. This approach allows the team and stakeholders to efficiently collaborate and craft the process to best suit actual ongoing needs, which leads to a faster, more accurate output.



BigR.io is a thought leader on Agile and hold Agile certification seminars; training the technical community in our approach to Agile development. Salient points of BigR.io's Agile methodology include:

- Short iterations of deliverable (tested) software we recommend a 1 to 3-week cycle.
- Transparency and stakeholder involvement:mutual decisions and no surprises.
- Definition of done: all delivered stories include design, development, testing, and documentation. Note: we create "just enough" documentation; balancing speed with artifact creation for continuity of knowledge.
- Relative and ongoing estimation utilizing story points and team velocity estimates to keep a firm grip on the reality of a project, rather than an idealized fiction.
- Modern testing automation code is built, tested, and deployed in a continuous and automated manner.

BigR.io delivers customized solutions and architectures that maximize results for clients - on time and on budget. We drive results with cutting-edge open-source and off-the-shelf technologies, pulling in the best tools and systems for each unique use case and environment.



We do not prescribe to a one-size-fits-all mindset, rather we examine currently adopted practices and adjust them where necessary. In general, for delivery methodology we place importance on:

- Gathering requirements that are represented as user stories. This is done in a collaborative manner with stakeholders such as product owners or the product manager. The user stories are written to fit into an iteration.
- The user stories are estimated using relative story points, and large "epics" are split into smaller, more manageable units. Planning sessions are conducted at the end of every sprint.
- Each story is assigned to an individual owner for task breakdown and focused management of story progress across contributors; implementation includes a brief design and a set of unit-tests to provide functional integrity and code coverage.



- As part of continuous integration, the stories undergo rigorous testing, and any resulting defects are addressed in the same iteration or moved to the next iteration based on priority.
- All unit tests are collected and turned into an automated smoke-test that runs on a nightly basis, or each time code is pushed to the integration branch.
- Peer reviews and a high level of team collaboration are encouraged at all times to maximize the probability of team success and delivery of a high-quality product.
- At the end of each iteration, completed stories are accepted by the product owner and marked as complete. Any deficiencies found during acceptance are addressed and corrected for future acceptance. Ongoing acceptance is encouraged where feasible. To be able to hit the ground running for the following sprint, the last day of every sprint is spent tackling all group meetings:
 - Review demo with signoff.
 - Retrospective finetune the process and approach.
 - Planning backlog grooming story assignments and high-level tasking.

All iterations or sprints proceed with non-functional requirements in mind. The BigR.io team lead assures that we never deviate away from attributes highlighted by these requirements. With the modern tools and techniques that we employ, there will be a boost to the "maintainability" of the system. With well-documented code and automated logging, we will have audit trails and granularity for monitoring quality.

BIGR.IO PROCESS

To address clients' requirements, BigR.io will first conduct a discovery process to gain a deeper understanding of the use case and operating constraints. Technology and infrastructure choices will be proposed to support the needed functionality, latency, cost, and other constraints. Where possible, expected trade-offs between architectural and technology choices will be presented.



BigR.io will present possible solutions to the client and work together to select the best option. Once a solution has been selected, BigR.io will outline a proof-of-concept (POC) to demonstrate general functionality and to help identify any risks or other points of concerns within the solution. This POC will also be used to help create an implementation plan to guide the production rollout of the selected architecture, including identifying re-development or integration of current applications into the new environment. If the POC reveals concerns, then an assessment of those concerns will be undertaken to verify the viability of the approach in a production setting.



Though each project delivery and time schedule depends on scope of work and team size, we generally prove out a Proof-of-Concept (POC) in 14-22 weeks. We break up this Phase 1 into Phase 1A & Phase 1B in order to separate out strategic alignment from tactical go-forward.

PHASE 1A

Discovery:

- Project kickoff and stakeholder engagement. Typically, a full day or two.
- Understand and capture the current-state technology and data landscape.
- Identify and quantify specific performance characteristics and reference use cases. Meetings will be conducted to gather this knowledge from existing users and other stakeholders.

Solutions Generation:

- Identify potential candidate solutions.
- Perform high-level feasibility assessment to identify finalist candidates.
- Deeper investigation of finalist candidates, identifying trade-offs and sensitivities.

Mutual Decision on POC scope and approach:

- Overview of finalist candidate architectures with tradeoffs.
- Detailed presentation of mutually-selected final solution architecture.
- Suggestion for "quick win" use cases to incorporate into POC.
- Implementation plan for POC.

PHASE 1B

Initial model design:

- First-round findings from the chosen candidate model(s).
- Completion of data pipelining.
- Template of all models generated.

Model training:

- Iteration of the primary predictions.
- Improvement on the secondary/tertiary model predictions.
- Demonstrate progression of prediction accuracy.

Conclusive findings concerning predictability of target problems:

- Report and present on the quality of the prediction model and future opportunities for machine learning.
- Demonstrate the ability to predict additional related classifications.

TIME SCHEDULE



Our experience shows that collaboration is the key to success. We collaborate closely with the startup's project and key decision makers to understand the details of the engagement. We utilize online communication tools effectively

to keep everyone on the same page. Regular informal and formal updates, brainstorming sessions, lunch and learn webinars, and workshops are conducted to create engagement and ownership in the deliverables. Agility is the key with tech squads, two-week sprints, demos, retrospectives and grooming initiatives. We stay ahead of the curve working towards your success.

Healthcare AI startups interested to partner with us can please email their non-confidential information to innovation@bigr.io

