squirro

Doing the Heavy Lifting – The Squirro-Added Value for Your Data Science Team

October 2019

Table of Contents

To Build or To Buy?

Almost every new technology that has emerged over the past few decades has presented to enterprises the same question – should we build, install, deploy, and maintain this technology ourselves or should we work with a partner organization to do it for us?

That question is as applicable to an on-premise CRM platform as it is to Software-as-a-Service and almost all points in between.

There is never really a definitive answer either. Whether through a carefully planned strategy or through good fortune with recruitment, some organizations will have the resources, budget, and talent to be able to build and maintain their own deployments, whereas other businesses will fall a long way short and buying in the expertise will be the smartest option. In some instances, building your own technology might be preferable, while in others, working with an external organization, an expert in that field, could be more advantageous.

Building your own technology will always have its proponents, but many enterprises are starting to come around to the idea that buying technology has its advantages. Recent IDC research, **Worldwide ICT Spending Guide: Industry and Company Size** suggested that software purchases will surpass app development in a few years. Furthermore, the most recent **CHAOS Report** that looks at software development, revealed that just 29% of in-house software development projects succeed.

This conundrum is also a significant one in the world of data science and Machine Learning (ML). If an organization has invested in its own data science team, is already building its own models that can process company data and, can utilize managed services, such as those offered by Google or Microsoft, what is the incentive to work with an external partner such as Squirro?

Squirro not only has one of the industry's most lauded Artificial Intelligence (AI) and ML platforms but also has a team of experts with a host of commercial experience in tailoring and productizing services. As a leading software company, Squirro has developed an open and scalable solution that can be the perfect vehicle for data science teams to bring their algorithms to their business users and drive ROI.

> **We believe that our proposition carries with it immense value for inhouse data science teams and can do much of the heavy lifting on their behalf.**

That's why this new whitepaper explores the challenges that organizations face when looking at technology adoption themselves and outline precisely the ways in which working with Squirro can benefit a company.

# The Challenges for Data Science Teams When Building Their Own Models

It's easy to see why some organizations might be drawn into building their own deployments. If they already have an on-site data science team, have established a data warehouse or data lake, and have the utilization expertise of the open source libraries that they need, from, say, Tensorflow to Keras to pandas. then why would they not want to manage everything themselves? Yet it's rarely as straightforward as that and there are a number of challenges to keep front of mind when data science teams look to build their own models.

A lack of resources

Even if a data science team is well-staffed with a selection of skilled employees, it doesn't mean that they can spend untold time on such projects, and it doesn't mean that they will have every base covered. ML requires specialized skills and knowledge and even the biggest inhouse teams may struggle to cover this.

It's also easy to under-estimate just how long ML projects can take. Doing everything yourself typically takes more resources than an organization usually has – gathering data, cleaning it, building the test and training sets, annotating and creating your training decks. Organizations typically allocate enough resource for the initial planning and prototype stage but then expect everything to run smoothly from the implementation onwards and do not always keep sufficient resource to manage this. That is when internal ML projects can often run aground.
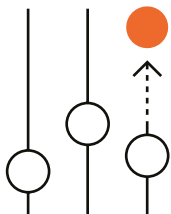
Ineffective productization

Turning a prototype into something that is more productized and of genuine use to the business is no small undertaking in itself. Even the best data science teams can lack a certain commercial know-how and what is developed in a lab environment can be a long way from what the finished product is like. This means that sometimes these projects get only so far before falling at the final hurdle, just before they can start to add true value to the business.
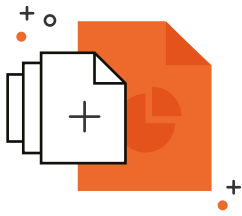
Using the full range of data sources

A key aspect to remember with ML is that it is only as effective as the data in which you put into it. If you include only certain data sources or are unable to work with all the relevant data, then it stands to reason that the results will not be as good. To be successful it requires an organization's own data - far more effective than an open source library – and the right methodology. An inhouse team working just with an open source library might not be able to do this.
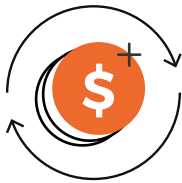
On-going costs

Using the inhouse team and existing resources can often seem like the most cost-effective approach and the one most likely to deliver strong Return on Investment (ROI). If everything runs according to plan and the team runs into no hitches, then perhaps this is true. But almost every IT adoption that has ever taken place has proven this not to be the case. This is especially true when considering that project requirements and scope tends to change over time.

When a data science team working with such large and varied data sets are trying to ask complicated questions of that data, the potential for things not to go as planned is even greater.

Maintaining a project, especially if it needs additional time and attention than was originally planned for, can be costly and any cost advantages that were envisaged from working inhouse are swiftly negated.

The Limitations of managed Cognitive Services

Many organizations seek a different option to working exclusively by themselves on their data science projects – deploying a managed service provider such as Google or Microsoft.

These are not without their benefits, but for an organization that is serious about getting the maximum value possible from its data, they do have many limitations.

While managed cognitive services in data science do have a convenience value, a major problem is that such services are primarily designed for problems trained with generic company unspecific data. Some providers have started to offer automatic machine learning, but they cannot really be used for machine learning or data science agnostic users - this is a significant challenge.

Furthermore, for many companies the usage of such cognitive services will most likely be limited to external or not sensitive data, as sending their own data into the public cloud is not desirable or possible for many businesses.

The Added Value from Squirro

Squirro provides a strong contrast with managed cognitive service providers which lack the flexibility required to really unlock the value in a data science project. Squirro supports all the state of the art ML frameworks and libraries; works with the most popular languages; and supports new and current models provided by these libraries. Other providers are much more limited in scope and lack the flexibility to work with such a variety of models.

Squirro Platform

Get the best from the Squirro AI Platform, Leverage ML Toolkits and Cognitive Services!
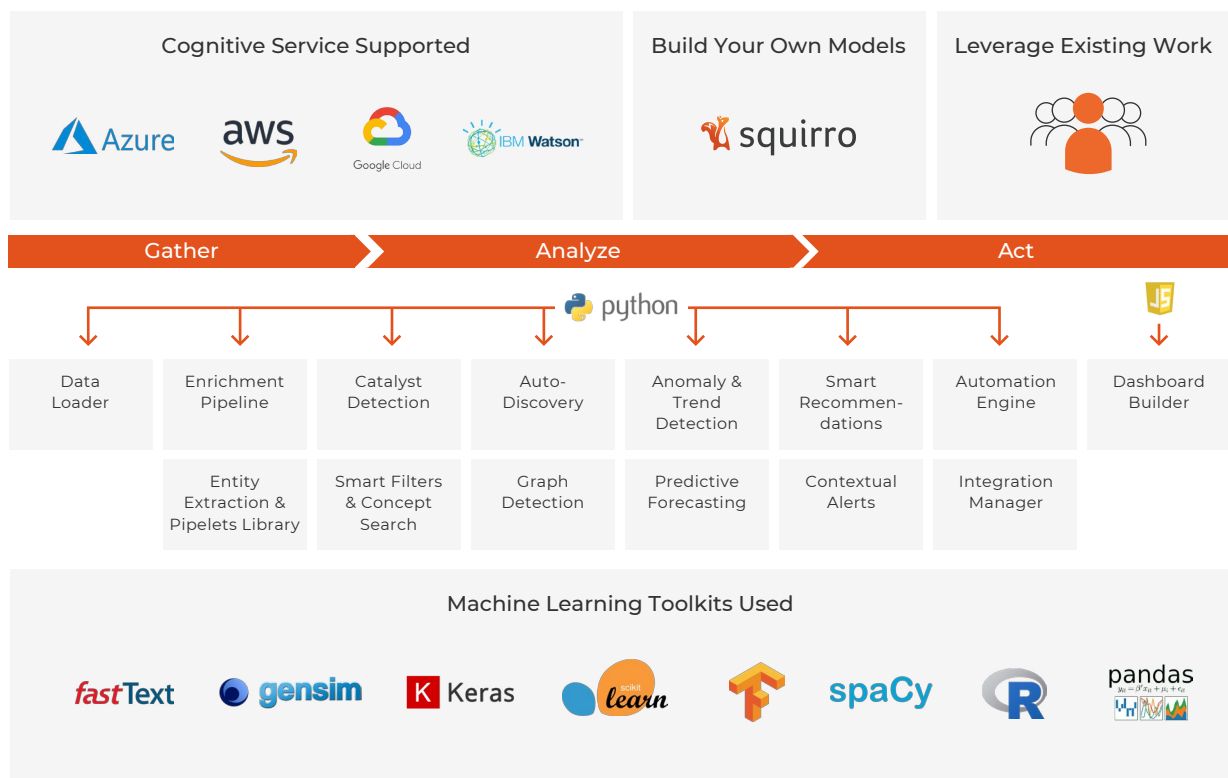
The Squirro platform offers a unique combination of state of the art solutions that span from big data to machine learning. The goal is to allow users to:

- Efficiently store, analyze and retrieve personal information.
- Efficiently apply state of the art analytics and machine learning.

All of that can be deployed on-premise to prevent data leakage and thus avoid any serious security issues. Squirro relies on ElasticSearch to store data efficiently. On top of that, many functionalities connected to Natural Language Processing are offered. With Squirro it is easy to apply NLP data transformation and extraction such as part of speech tagging or keywords extraction.

Another key piece of the Squirro-platform is the Machine Learning Service. Squirro offers MLOps capabilities to train, deploy and maintain ML models very efficiently and with minimal effort. Squirro relies on cutting edge data encoding strategies such as Bert [1] or Doc2Vec [2] that transform the document in a 'machine-readable' format by also considering information such as meaning and semantics of the input. Efficient encoding is fundamental in to be very effective in Machine Learning.

Afterward, in Squirro two machine learning strategies can be applied:

- Unsupervised learning: we rely on state of the art cluster algorithms such as DBSCAN [3] to group items together or to find outliers.
- Supervised learning: a full range of supervised learning strategies are available in Squirro: they span from traditional methods (Random Forest, …) to Deep Learning (LSTM [4], …)

Moreover, Squirro offers MLOps capabilities. Once a model is trained with Squirro it takes two simple clicks to deploy it within a different Squirro instance. Another fundamental element is model maintenance. It is very common that models degenerate over time, a process known as Conceptual Drift. To avoid that, Squirro uses a set of solutions that range from drift identification [5] to drift prevention with constants and automatic retraining.

Store   Analyze   Retrieve   Analytics   Machine learning

[1] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[2] Le, Quoc, and Tomas Mikolov. "Distributed representations of sentences and documents." International conference on machine learning. 2014

[3] Ester, Martin, et al. "A density-based algorithm for discovering clusters in large spatial databases with noise." Kdd. Vol. 96. No. 34. 1996.

[4] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

Currently, Squirro is extending its ML capabilities in different areas. For instance, a new concept of ensemble strategy is proposed. Ensemble in ML consists of executing simultaneously different ML algorithms and selectively filtering the results to improve the quality of the inference. This solution is known to have serious requirements in terms of computation, but Squirro is developing a selective ensemble approach able to leverage different methods only when they are really necessary. This will offer the quality of ensemble approaches without computational resource abuse.

The whole Squirro ML service is completely extendible and can extend the framework within your own environment. So, if any organization already has a data science lab, and comes up with something that works in the lab, Squirro can help greatly, uplifting the prototype into a real production use case.

As a private and secure environment, Squirro can also be deployed on premise without the need to give away data. For any organization concerned about security when sending data to a public cloud provider, this an important distinction.

"

Finally, almost every IT project is comprised of 80% glue code, custom-written programming that does not offer much in terms of functionality but binds together different parts of code that otherwise would not be compatible.

This 80% glue code is written for many different use cases, so organizations can use Squirro to collect all the data to add more interesting and bespoke pieces on top.

At Squirro we embrace and support the usage of our solution by internal data science teams and have enabled many of them already at our clients through our Squirro Certifications. We are convinced that Squirro can be the perfect vehicle to help data science teams to drive their AI & ML strategy to success through the usage of our out of the box applications, or by leveraging our modular and scalable solution to bring insights to all your business users.

[5] Gama, João, et al. "A survey on concept drift adaptation." ACM computing surveys (CSUR) 46.4 (2014): 44.

Conclusion

Squirro is a provider that is easily and quickly embedded into core solutions like CRM, CSM, ITSM and many others, visualized, and therefore brought into a business context.

It brings great value to a company and instead of data science teams failing because they are over-stretched and lack the tools to do the job effectively, Squirro does the heavy lifting that ensures they can be one of the most valued and appreciated teams in the entire business.

Organizations are aware of the need to get more value from their data and know that their data science teams need help to extract that value. Squirro's Insights Engine and Augmented Intelligence Applications are among the best in the industry and are deployed by many data science teams in a variety of industries and different locations.

For more details on how Squirro can help your organization's data science team to truly deliver, then please get in touch with us.

Squirro – © 2019 Nektoon AG   **www.squirro.com**   **contact@squirro.com**
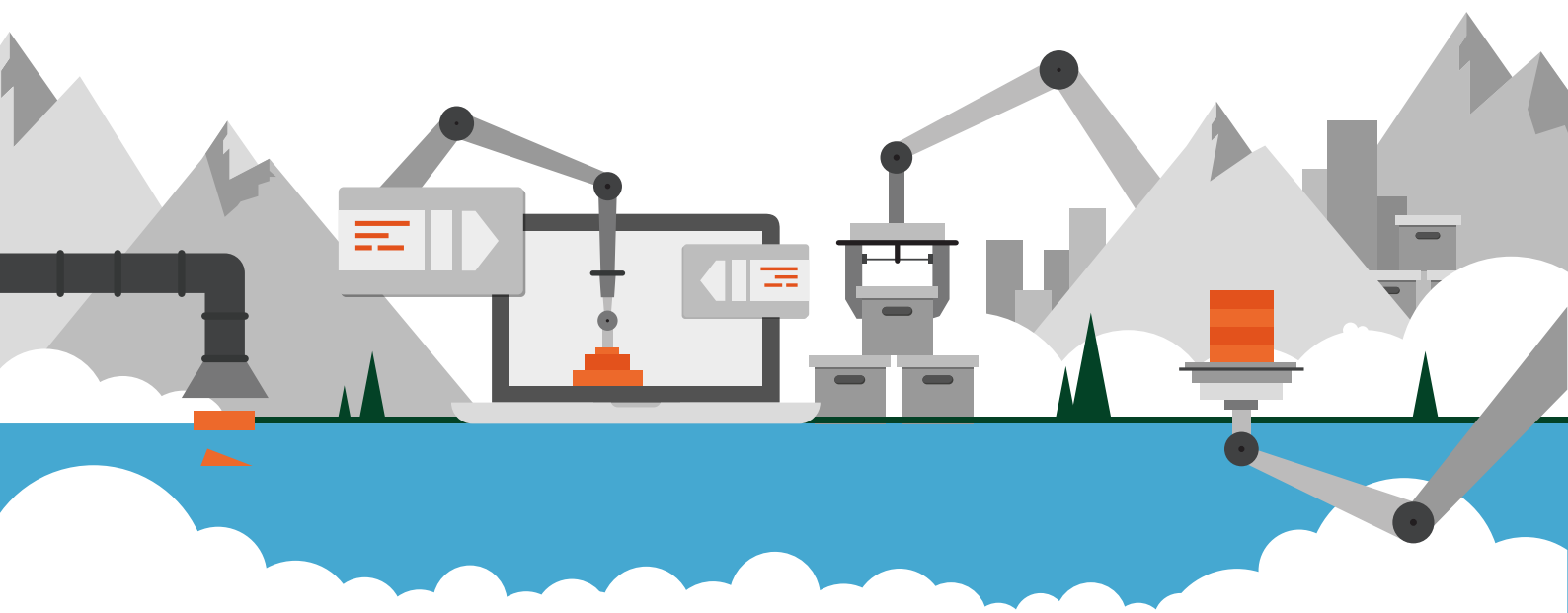
About Squirro

Businesses generate and convert new leads, improve existing relationships, and optimize processes using Squirro's vertical-specific Augmented Intelligence Apps, which combine human intelligence with powerful AI. An Insights Engine at its core, Squirro applies cognitive search, advanced analytics, and intuitive dashboarding to unstructured data to provide new opportunities, next-best-action recommendations, and real-time 360° client cockpits. All which can be integrated within CRM Systems, Core Banking Systems, or Insurance Solutions.

Squirro works with global organizations, primarily in the Financial Services, Insurance, Telecommunications, and Manufacturing industries. Customers include Brookson, Investec, Helvetia Insurance, and Axis Capital. Founded in 2012, Squirro is currently present in Zurich, London, New York, and Singapore. Further information about AI-driven business insights can be found at **www.squirro.com**

Squirro Switzerland

Badenerstrasse 120,
CH-8004 Zurich,
Switzerland

+41 44 586 98 98

Squirro Americas Inc.

33 Irving Pl.
New York, NY 10003
USA

(929) 999 2173

Squirro Deutschland GmbH

20 Oskar-von-Miller-Ring
80333 Munich,
Germany

+49 89 2154 7429 0

Squirro United Kingdom Ltd.

1 Fore Street
London EC2Y 9DT
United Kingdom

+44 333 939 9973

Squirro Asia PTE. LTD.

190 Middle Road #16-01
Fortune Centre
Singapore 188979

squirro