

analytics 1305

www.analytics1305.com
www.fast-lab.org

Senior Team

Dr. Alexander Gray, CEO.
Dr. Nikolaos Vasiloglou,
CTO and CFO.
Abhimanyu Aditya.
Dongryeol Lee.
Ryan Riegel.

Location

Atlanta, GA.

Industry

Analytics.

Products

1) The first back-end software capable of advanced machine learning analytics on massive datasets, 2) Custom high-performance modeling for business analytics applications.

Company Type

Privately-held start-up.

Intellectual Assets

Proprietary massive-dataset machine learning methodology and commercial software. Maintainers and distributors of first scalable machine learning library, open source.

MISSION: MASSIVE ANALYTICS

Analytics 1305 is the *massive data analytics company* – the only source for state-of-the-art machine learning methods which scale to very large and complex datasets. We license our first-of-a-kind scalable advanced analytics back-end software, and also develop custom statistical modeling and optimization tools. In a nutshell:

- **Advanced methods on massive data.** We have the only technology which can perform a large array of advanced analytics (state-of-the-art machine learning/data mining/optimization - not just simple queries and methods) on massive datasets.
- **Large-scale systems.** Our software is being integrated into the pipelines of the world's largest-scale scientific projects (including the Large Synoptic Survey Telescope and the Large Hadron Collider), supermarket retail forecasting systems, automatic real-time spam blacklisting engines, document and image search engines, and first-of-a-kind cancer diagnosis systems.
- **Very fast, comprehensive software toolkit.** We license our powerful library (containing over 100 data analysis methods), which utilizes a combination of algorithmic advances, mathematical acceleration techniques, parallel/multicore computing, and efficient coding; we also develop custom application-specific methods.
- **World-class expertise.** Our sister organization is a large academic lab of 25 people, the FASTlab at Georgia Tech, which connects us to the very latest state-of-the-art in the research literature as well as a network of top experts across the nation and internationally; our work has won or been nominated for a number of awards.

Projects. Analytics 1305 brings a long history of experience in analyzing massive and complex datasets and developing custom solutions since 1993 (over 16 years), in over 50 combined projects across science (e.g. astronomy, biology, chemistry, geology), engineering (e.g. network security, remote sensing, audio fingerprinting, medical diagnosis, document search), and business domains (retail forecasting and promotion planning, CRM, options trading, online advertising). Our software is being implemented in the pipelines both of the next-generation massive-scale astronomy instruments, which will collect time series of billions of sky objects per month, as well as the Large Hadron Collider, which will collect a stream of millions of measurements per second. Other applications include automatic spam blacklisting based on the temporal email sending patterns of hosts using 300 million emails per day, and sales forecasting for a major supermarket's massive product inventory.

Company. Five colleagues at the Georgia Institute of Technology founded Analytics 1305, a privately-held company, in March of 2009. The company was seeded by a large contract to provide the data analysis back end for a software-as-a-service business intelligence/retail forecasting product, and is

developing new partnerships with companies ranging from small startups to the very largest. Analytics 1305 considers relationships ranging from consulting to revenue sharing.

Bios. The members of Analytics 1305 bring industrial experience in both large-scale data mining / optimization and advanced software development at Georgia Tech, NASA, Google, Microsoft, and Trilogy. Dr. Alexander Gray received his bachelor's degrees in Applied Mathematics and Computer Science from the University of California, Berkeley and a PhD in Computer Science from Carnegie Mellon University. He worked for six years at NASA's Jet Propulsion Laboratory in its Machine Learning Systems Group and is now a professor in the College of Computing at the Georgia Institute of Technology. He is director of the FASTlab. His work has been described in staff-written articles in Science and Nature, including contributions to work selected by Science as the Top Scientific Breakthrough of 2003. He has won best paper awards in statistics and data mining and the National Science Foundation CAREER Award. He gives invited tutorial lectures on massive-scale data analysis at the top data analysis research conferences, government agencies, and corporations. Dr. Nikolaos Vasiloglou received his BSE in Electrical Engineering from National Technical University of Athens and his PhD from the Georgia Institute of Technology. He has worked in several Silicon Valley companies, such as HP Labs, Google and Admob. He has experience in large scale optimization, manifold learning and advanced programming.

APPROACH: CONNECTION TO STATE-OF-THE-ART RESEARCH

Analytics 1305 takes a unique approach which brings to bear the intellectual strengths of the academic research community upon the extremely difficult problems arising in analyzing massive and complex datasets. While unusual, this approach lends an unmatched sophistication to our capabilities, relative to other competitors in the space of analytics companies.

Our research lab. Analytics 1305 maintains a close link with its sister organization, the Georgia Tech FASTlab, (Fundamental Algorithmic and Statistical Tools Laboratory, www.fast-lab.org) which is the premier academic lab devoted to the problem of massive-dataset machine learning and has developed and continues to develop the most efficient and effective algorithms for many fundamental problems in statistical data analysis, including state-of-the-art approaches for difficult optimizations and fast evaluation of complex models. The FASTlab is directed by Prof. Alexander Gray and consists of approximately 25 top-notch graduate students and post-docs, one of the largest labs in Georgia Tech's College of Computing. Hot-off-the-presses techniques developed in the lab are transferred immediately to Analytics 1305 on an ongoing basis, and the lab works on difficult research problems inspired by industrial problems encountered by Analytics 1305. In other words, the FASTlab is the large research arm of Analytics 1305, and Analytics 1305 commercializes FASTlab's technology. The company is named after the lab's room number.

Connection to the research community. Analytics 1305 stays abreast of the very latest developments in the large and diverse machine learning/data mining/statistics research community through participation and publication in conferences and journals. It is using its relationships in the field to develop a unique research commercialization "co-op" system, in which selected authors of the latest effective techniques earn royalties on a per-algorithm basis by having their methods incorporated into our libraries, leveraging the economy of scale afforded by the company to allow individual PhD students, professors, and other researchers to benefit from their ideas where it is not likely to otherwise be possible. In this way, Analytics 1305's data analysis technology is unmatched in comprehensiveness and state-of-the-art sophistication by any other organization, either commercial or academic. At the same time, Analytics 1305 gives back to its roots, the research community.

Pioneering MLPACK software. Analytics 1305 maintains and distributes MLPACK, the first comprehensive and scalable machine learning library, the analog to linear algebra's LAPACK. MLPACK is distributed as open source under the GPL 2 license, which is generally viable only for non-commercial use. This is performed as a service to the scientific and research community, creating a strong conduit for intellectual interchange and advancement which Analytics 1305 draws upon to create the most advanced data analysis solutions currently possible. MLPACK is designed only for datasets which fit in RAM. Analytics 1305 commercially licenses its much-larger-than-RAM technology, MLPACK Ultra, which also contains a large number of further methods, both published and unpublished.

CAPABILITIES: SCALABLE ADVANCED ANALYTICS

Scalable analytics. Analytics 1305 has the most capable and powerful technology for analyzing large and complex datasets of any organization. It combines state-of-the-art technologies based on statistics/data mining/machine learning, efficient algorithms/data structures, and parallel/distributed/cloud computing – including both published and unpublished concepts. Other approaches to massive-dataset machine learning rely on inefficient straightforward parallelization schemes such as Map-Reduce/Hadoop or stochastic gradient methods, neither of which are effective for complex machine learning methods which cannot be treated as simple linear aggregations. *Ours is the only available technology which scales state-of-the-art methods (those with the highest statistical/predictive accuracy) to massive datasets.* We can perform advanced machine learning operations in seconds that would otherwise require days or even years with other existing technologies such as in R, SAS, Matlab, Weka, or other software.

Advanced analytics. For many, “analytics” consists of simple SQL queries. While our capabilities include the ability to perform such simple database computations scalably, we have developed the unique capability of performing *advanced* prediction and modeling methods on massive datasets. Analytics 1305's software either currently contains or will very soon contain the fastest known algorithms for the most powerful and useful machine learning methods including principal component analysis, nearest neighbor, support vector machines, manifold learning, kernel estimators, kernel machines based on semi-definite programming, nonparametric Bayesian classification, Gaussian process regression, hidden Markov models, Kalman filters, autoregressive time series models, independent component analysis, non-negative matrix factorization, L_1 and L_0 regularization approaches, hierarchical clustering, nonparametric mean-shift clustering, mixture models, decision trees, spatial statistics, and discrete choice models, and the list is quickly growing. Our software also includes state-of-the-art optimization capabilities, including unconstrained optimization, mixed integer linear programming, quadratic programming, semidefinite programming, and stochastic programming, and advanced data visualization capabilities. We also develop custom methods to treat difficult real-world aspects of clients' data or other requirements.

Products. Analytics 1305 develops and licenses 1) machine learning/analytics libraries as highly scalable back ends for customers' front end products, and 2) custom scalable and highly-accurate machine learning/analytics methods for clients' needs, which may be new products or internal applications. Our software can be configured for existing computer clusters, clouds, or single machines. In addition to raw speed and memory efficiency, the software is designed to be easily integrated with a variety of platforms and languages, and is highly portable. We are also developing a standalone visual front-end for scalable advanced analytics, which will be available in 2010. Our extensive market research on today's available analytics products is available to our collaborators and partners.