Data Scientist

What we do

We help governments, intelligence agencies and corporations find solutions to their data and analysis problems. Our software is delivered worldwide via our partnership with IBM and twenty other sales partners.

Our flagship software product, Sintelix, is an advanced text analytics engine and is currently one of the fastest, most accurate and highly configurable unstructured data processing systems on the market. The company is enjoying a period of rapid sales growth at well above 100% a year.

A recent evaluation by the US Army said that there are two companies in the world that have Sintelix's capability; the other one is Google.

The Role

We are looking for one or two gifted and highly motivated Data Scientists to drive the creation of the advanced algorithms which are the life blood of the company's software.

The Data Scientist(s) will join the existing team responsible for developing algorithms for Sintelix, Sensuris and Inventris. We are looking to augment the current team's strengths in optimisation and discrete decision making.

They will apply their experience of success in research and development to create robust, high quality solutions that provide an excellent performance and customer experience. Their skill in Java will enable them to create algorithms (which after some clean-up) can be used directly in production software.

Most developments will involve (i) surveying, finding, evaluating and adopting existing technology (ii) expanding, extending and constraining existing algorithms, (iii) developing new algorithms and (iv) iteratively improving algorithms to achieve performance levels required by the market.

Supervision will be provided by senior staff with PhDs and relevant domain knowledge.

Other responsibilities include:

- Clarification and elaboration of requirements from managers and customers;
- Communication results and ideas to customers and managers in the form of presentations and well written scientific reports;
- Careful evaluation of performance and robustness of solutions; and

- Challenge and expanding thinking within the company.

Data Science at Sintelix

Over the last decade Sintelix and its collocated sister company Semantic Sciences has pioneered a number of high-performance algorithms that are now embedded within the companies' products; they include:

- **Entity Extraction in English** at F1 score of greater than 95%

- **Geocoding** – more than six times higher F1 score than the market leader

- **Entity Resolution** – rapid, high performance (best score of any commercial tool on the "John Smiths" corpus)

- **Community Structure Analysis** – very flexible algorithm that generalises the concept of community structure analysis

- **Document Classification** – robust, high performance algorithm

- **Optimal Resource Allocation** – applied to selection of reviewers for research proposals

The algorithms we develop have to work well both in terms of objective capability and customer satisfaction. We therefore seek to design algorithms that are:

- *Accurate* – with low error and miss rates
- *Fast* – both to configure/learn and to run
- *Scalable* – in terms of data and processing
- *Fixable* – there is an easy and cheap performance improvement process
- *Configurable* – the algorithm can easily be configured to work well in its application
- *Productive* – the process of using the algorithm requires low effort and skill
- *Deployable in a JVM environment* – to make deployment within our software flexible and cheap

Specific Skills Required

- Knowledge of some and preferably all of the following domains:
    - Machine Learning: including pattern recognition, deep learning, MCMC, inference on Markov chains
    - Computational Statistics: including classification, regression, clustering, time series analysis and prediction, statistical modelling, Bayesian statistics
    - Mathematics: especially algebra, matrix algebra, scalar, vector and matrix differentiation
    - Computational methods: especially optimisation
    - Language Technology/Computational Linguistics: especially surface parsing
    - Data cleaning, record linkage and master data management
- Successful experience in developing and improving algorithms in at least one area within Machine Learning, Computational Statistics, Language Technology and Record Linkage.
- Rapid assimilation of research literature within a topic under investigation.
- Good knowledge of the current state of the art in your area of expertise.
- Good programming ability – preferably in Java.

Qualities we are looking for

We are looking for candidates who are …

- Warm, genuine
- Effective
- Highly intelligent and pick up complicated concepts quickly
- Can rapidly and thoroughly understand a complex software product
- Able to synthesise a consistent and integrated understanding of the overall mission
- Creative and can inspire creativity
- Strategic and analytical … can join up the dots
- Great to work with
- Responsive and careful
- Excellent communicators orally and on paper with the ability to
    - write accurate, grammatical and persuasive text
    - create rapidly scannable materials
    - achieve the communication purpose and
    - create materials that are elegant and have high impact

and

- Have an eye for detail
- Are well organised and
- Produce on time!

Working at Sintelix

At Sintelix, our goal is to create a space where our people have the chance to achieve excellence within a warm an encouraging environment. We go to great lengths to create wonderful products and provide excellent service to our clients. Our solutions are usually world-leading in quality and performance. Elegance, simplicity and effectiveness flow into everything we do.

We take care to base all our decisions on evident truth and consideration for others. In the atmosphere of trust that pervades the company, we can learn from each other and celebrate each other's development and contributions.

We offer flexible working hours and a bonus system built around sharing the company's success.

Interested in applying

If you would like to discover more about this opportunity or apply, please contact Mitchell Hodgkinson at jobs@sintelix.com. Include a current CV and letter of introduction. You may be asked to perform some skills tests and provide references as part of the recruitment process and your interest will be treated in the strictest of confidence.