# Job Description:

Senior Consultant in EXL Service, Product Development Lab, a team of techno-analytics enthusiasts to build data applications - solutions or products.

My focus areas are:
• Text analytics (unstructured text extraction and classification, machine comprehension, e-mail classification and extraction using graphs and deep learning, NLP, NLU)
• Image analytics (OCR, Image processing, segmentation,object detection, visual search and classification using Deep Learning )
• Speech analytics (speech to text transcription using Deep Learning, NLP, emotion detection and speaker diarization using ML, agent performance monitoring)
• Smart meter analytics - IoT data (moving k-means, segmentation, time-series demand forecasting)

I design, develop and implement wide range of deep learning algorithms using Python/Spark/Tensorflow/PyTorch libraries and package the application in sleek Dockers. My responsibilities are end-to-end: research, design, development and collaboration. I have also been a key member of client/investor capability showcase, running recruitment from premier institutes and collaborating on industry engagement activities e.g. NASSCOM summits, ODSC summits etc.

I believe learning is progressive and active learning via competitions, knowledge exchange at meet-ups, summits, conferences etc. and MOOCs adds experience to build upon.

Latest Binge : Language Modeling (Exploring ULMFiT and BERT)

| **Job Duty** | **Percentage of time spent per week on this duty** | **Detailed description of this job duty in simplified terms.** |
|---|---|---|
| 1.<br><br>**Understanding and Exploring Data** | 15% | Our clients come to us with a lot of data but they do not understand how to read or visualize and monetize it. So as a primary objective, I understand the data first before getting into formulating or solving any queries posted by clients. I use my skills in statistics, programming and data management in :<br><br>1) Providing a transparent structure to data<br>Ex: *ordering it into relevant categories*<br>2) Filtering all but data of interest<br>Ex: *removing duplicates vs irrelevant, fixing structural errors, handling missing data, fixing outliers etc.*<br>3) Mining by applying methods from statistics to recognize patterns and place data in mathematical context<br>Ex: *feature engineering, feature selection etc.*<br>4) Representing relevant data using a basic visual model<br>Ex: *bar graph, list, or tree etc.*<br>5) Refining basic representation of data to make it more visually engaging<br>6) Interact with fellow workers to define scope of the data<br><br>This is a repetitive exercise and I keep iterating over versions of data, from which we choose the best version that serves the problem statement. |

| | | |
|---|---|---|
| 2. **Formulating a Problem Statement** | 10% | In order to serve the needs of clients, I formulate problem statement. I create a flowchart and graphical representations, which breaks down the problem explaining sub-tasks involved at each stage in solving the problem. I do so, after a string of iterative brain-storming sessions with members of Technical (IT) support, Data Engineers, Data Scientists and Developers.<br>We engage in workshops were we<br>• understand the scope of problem,<br>• underline questions in the problem,<br>• discuss various alternative theories and options and<br>• create process maps to help management team visualize the complexities involved.<br><br>We communicate this to clients by creating a business requirements document. |
| 3.<br>**Consulting and Research** | 5% | I offer subject-matter expertise to the team in evaluating the value of application of machine learning and advanced analytics interventions to various projects. To support my arguments, I perform *feasibility, logic and randomness study* along with *cost benefit analysis* with both technical and business factors into consideration.<br><br>I also get involved in active research concurrently, to browse through available literature and get acquainted with the selected problem. I try to understand the problem thoroughly by rephrasing the problem in possible different ways in meaningful terms from an analytical point of view. At this stage the different versions of data, created earlier helps brew different ideas.<br>For Example: I develop a working hypotheses i.e., making tentative assumption in order to draw out and test logical or empirical consequences of methodologies that could possibly answer the problem.<br><br>During the course of research, I create a list of methodologies that can possibly address the problem and their complexities.<br><br>For the purpose of creating reports I use *LaTex* to maintain industry standards in report writing. |
| 4.<br>**Defining / Selecting Methodology** | 20% | At this stage I have,<br>• explored the data<br>• valued opinions of experts in IT, Data Engineering and Development<br>• prioritized items from management team<br>• a clear problem statement<br>Added to the above, is the research that went into selecting a suitable candidate list of methodologies to solve the problem.<br><br>Keeping in mind the requirements put forth by various teams, I take up a sample dataset from the original data provided. Here, I use my expertise in statistics to come up with a sampling technique that is most suited for the data, *viz., Deliberate sampling, Simple Random sampling, Systematic sampling, Stratified sampling, Quota sampling, Cluster sampling, Multi-stage sampling, Sequential sampling etc.,* and choose a sample that is a representative of the original data.<br><br>I train models (*statistical, Machine Learning, Deep Learning etc.*) that are suitable for the problem thanks to the research and subject-matter expertise. At this juncture, I also indulge in *feature selection or engineering* which itself is a big area of research, for certain techniques when required. With the given pre-defined standards of accuracy, I compare the models and choose the best one that |

| | | |
|---|---|---|
| | | satisfies the constraints. If the models do not respond to reciprocate the desired values, I then attend to devising suitable methods either a new discovery or an ensemble of various existing algorithms to tackle this problem.<br><br>For Example:<br>One method of dealing with the non-response problem is to make a list of the non-respondents and take a small sub-sample of them, and then with the help of research and experts' vigorous efforts can be made for securing response.<br><br>Once we secure the desired accuracy, I am now concerned about the computational constraints listed by the client's technology team with respect to the infra-structure they possess. I attend to it, by getting back to the design of the algorithm to find avenues of fitting in complexity reductions, using techniques like *dimensionality reduction, shared weights for all the layers of neural networks, freezing/unfreezing of layers during back propagation, batch processing, optimization algorithms etc*. I use customized python codes to carry-out the above mentioned.<br><br>After locking all the metrics in methodology, I now start preparing the necessary reports and coding modules that fits into a developer's framework, to train the selected algorithm on the original dataset either in client's environment or EXL's native machines or AWS instances depending on the mandate.<br><br>For this purpose I use *Machine Learning algorithms* like *SVM, Naïve Bayes, Decision Trees etc., Deep Learning algorithms* like *Neural Networks, MLP, CNN, RNN (GRU, LSTM), Attention Mechanisms, Auto-Encoders Decoders, GANs etc., NLP concepts* like *Bag of Words, Tf-idf, Named Entity Recognition, Semantic Parsing, Linguistic Parsing etc., NLU concepts* like *Machine Comprehension etc., Signal Processing concepts* like *Acoustic Mining, Time Series analysis etc.* amongst other *Optimization Techniques and Genetic Algorithms.* |
| 5.<br><br>**Coding the module for selected methodology** | 10% | I take up the responsibility of coding relevant deep learning and machine learning modules. Coding for algorithms that I have designed, helps me tweak and fine tune parameter at will, during training. Coding my own architecture also make it easier to formulate the appropriate optimization algorithm for the purpose as training a model is an iterative process. I finalize the best model appropriate for deployment into production and my modules seamlessly integrates into the developers' framework.<br><br>Apart from building modules for an application, I also have a sound interest to program *Spark modules* to leverage the distributed computing capabilities and reduce the computational complexities while training data intensive and very deep neural models.<br><br>For Example: I conceptualized the *Machine Comprehension algorithm* and wanted to test it. I searched the web for publicly available trusted datasets and code my algorithm from scratch. This assisted me in underlining the flaws in the architecture and understand the computational limitations to train the model. I had to use a GPU in-order to train it else it would have taken months to train on a CPU based framework.<br><br>Coding also helps me decide on the trade-offs that can be incorporated within the algorithmic design and come up empirical approximations where-ever needed without affecting the overall accuracy. |

| | | |
|---|---|---|
| | | For this purpose I use coding in languages like *C, C++ , Java, Python, R, Matlab* and use frameworks like *PySpark, Tensorflow, PyTorch, Keras, Theano etc.* |
| 6.<br><br>**Experimenting with Alternate theories for Comparative Study** | 15% | I conduct experiments to test alternate theories/algorithm that could potentially solve the problem from the list of options we had prior to methodology selection. I then create a detailed theoretical report supported by results and insights from computer-based experiments to generate and confirm or confront conjectures.<br><br>In the details report I specify the<br><br>• new patterns and relationships<br>• use graphical display to suggest the underlying mathematical principles<br>• reasoning to falsify conjectures with mathematical proofs<br>• suggesting new synthetic-hybrid approaches that could increase the accuracy<br>• confirming analytically derived results<br><br>This helps me in creating a similarity/dissimilarity charts for comparative study amongst the different methodologies in terms of certain standard statistical KPIs. By this, I also engage in talks with team members and interested parties on the trade-offs between computational speed and algorithmic accuracy. |
| 7.<br>**Error Analysis** | 15% | I usually build a quick model for any problem and iteratively identify errors and keep fixing them. This not only helps in improving the accuracy of the model but also gives me a new perspective of visualizing the problem.<br><br>Since most of the time, I deal with complex supervised deep learning architectures whose base is a neural network, I categorize the errors into *reducible and irreducible*. I come up with fixes to improve the model's performance by decreasing the reducible error. I further build a methodology to classify reducible error into *bias and variance* which depend on human errors involved and sampling effects respectively.<br><br>For Example:<br>In supervised learning, I reduce the available labeled dataset into three broad categories a) train b) validation c) test. The training set is used to teach the model to recognize and learn various features and parameters to produce predictions. The dev set is used to optimize the trained model so as to maximize the accuracy of the model. The test set is used to run the optimized model and see its performance in the real world.<br><br>I then re-train the model after sorting the effects of incorrectly labeled data and mis-matched distribution over train, validation and test dataset.<br><br>Error analysis is done at every stage whenever a new model is developed during methodology selection as well as experimentation phase. |
| 8.<br><br>**Monitoring performance and Initiating Feedback loop** | 5% | With the applications now in ready to go for production environment, I get involved in rigorous testing of the software/application with the QA team. I assist them in simulating practical real-world scenarios and monitor the performance of our application. If the system that we designed has a human-in-the-loop dependency, I also co-ordinate and guide the operation's team of EXL at various |

| | | |
|---|---|---|
| | | locations. I initiate a feedback loop between the agents in operation's team and the testing team to produce a fail-safe application.<br><br>I also keep track of the impact of the application with respect to the user by monitoring few quantitative KPI such as reduction in *Average Handling Time of the process, improvement in quality of output, reduction in rework by the operators etc.* These given feedback are tracked and models are updated periodically. |
| 9.<br>**Talks/Conferences and Active Research** | 5% | I share the knowledge that I possess over a certain subject-matter with a wider community via talks in universities, conferences, meet-ups and EXL town-halls. I strongly believe in learning via interaction and discussions.<br><br>I am guest lecturer at Netaji Subhas University of Technology, Delhi on the topics of Machine Learning, Deep Learning in NLP. I attend conferences and engage in talks with senior researchers in the industry to spread and obtain new research in the field of mathematics and computational sciences. I organize meet-ups within operational teams in service industry to drive the concept of democratization of AI based application in day-to-day professional endeavors.<br><br>I keep learning by enrolling myself in various MOOC platforms on various subjects across different levels in application development and problem solving. I also draft detailed reports on the business case studies that I have come across working on various projects as a potential subject of active research in the field of applied mathematics. |