**Abstract:**
LinkedIn Data Scientists crunch through petabytes of data and build models ranging from suggesting your next connection, connecting to your next employer, and finding whether your profile is fake. In this talk, I will cover the end-to-end process of building, deploying, and validating a model at LinkedIn, and how it is different from similar work in an academic context. In particular, I will cover the importance of knowing the domain and feature extraction. I will also talk about how to work with big data and the new challenges it brings to the table, and the analytics stack (Hadoop, Hive, Pig, R, Azkaban, etc) that we use to solve those challenges. I will talk in detail on some of the problems that my team (Security) faces, especially developing models in an adversarial role. Finally, I will take time to answer the question that many people ask me: Who is a Data Scientist and how can I become one?

**Bio:**
A Columbia alumni, working as a data scientist in LinkedIn’s Security team, where his first priority is keeping the site secure against fraud and abuse by malicious users and botnets. He also helps in keeping customer data secure in big data platforms like Hadoop. For more information have a look at his LinkedIn profile [www.linkedin.com/in/karthikmramasamy/](http://www.linkedin.com/in/karthikmramasamy/).