

# Ratnakar Pawar

## Consulting Data Scientist - Cyber Warrior Network

Chicago, IL - Email me on Indeed: [indeed.com/r/Ratnakar-Pawar/e551e0f20e0c82a6](https://www.indeed.com/r/Ratnakar-Pawar/e551e0f20e0c82a6)

A former data engineer with recently acquired machine learning/data science skills, the ability to architect machine learning and analytics pipelines from the ground up, and an unquenchable curiosity. Able to build prototypes quickly and leverage a moderate dose of applied statistics and mathematics.  
Willing to relocate: Anywhere  
Authorized to work in the US for any employer

### WORK EXPERIENCE

#### Consulting Data Scientist

Cyber Warrior Network - May 2017 to Present

Helped building model to identify similar jobs for armed services veterans. To ingest resumes, cleanup data and extract useful features from corpus. nltk, mysql, pandas

Uptake - May 2017 to August 2017

Worked on a IOT security problem, for a water treatment facility data. Identify attack signature and classify the event as attack or non-attack. xgboost, clustering, timeseries, component analysis, pandas, sklearn

#### Analytics Engineer

Citicorp - February 2015 to August 2016

Design and build production-ready data pipeline to fetch events from various applications to a central data store. Lead the solution onboarding for internal stakeholders and handover to the production teams. Including the documentation and internal demos to multiple teams. Hadoop, flume, hive, log4j, syslogdm disaster recovery, Spark

#### Data Engineer

Telecom SME - January 2010 to February 2015

Designed and deployed TB Scale secure data solutions using Hadoop ecosystem products. Worked profoundly on creating on-demand Hadoop cloud for development environments using virtualization solutions like Docker and

vagrant. Developed integrations between the bigdata platform and proprietary telco analytics products. Tested and developed integration between Amdocs products. Involving scalability tests on distributed middleware platforms.

Analyzing and debugging applications/systems, part of product SDLC and regression testing. hbase, puppet, docker,

distributed clusters, weblogic, websphere, oracle, Kerberos, Hadoop, MapReduce, vagrant

### Honors and Awards

R&D president's award for contribution towards success of Amdocs product portfolio. Featured in top 20 entries (from over ~1000 entries) for the product security hackathon. Consistently won excellent contributor award from 2010 to 2014 at Amdocs. Praise award for design and implementation of log analytics pipeline at Citi.

Talks, Professional

## EDUCATION

### **Master of Data Science in Data Science**

Illinois Institute of Technology - Chicago, IL

2016 to December 2017

### **Master of Computer Science in Bachelor of Computer Science**

University of Pune - Pune, Maharashtra

2012 to 2015

## SKILLS

Hadoop (6 years), HADOOP (6 years), APACHE HADOOP MAPREDUCE (5 years), APACHE HBASE (5 years), ECOSYSTEM (5 years)

## LINKS

<http://github.com/rpwr021>

<http://kaggle.com/rpwr021>

<http://linkedin.com/in/ratnakarpawar>

## ADDITIONAL INFORMATION

### Technical Skills

Machine Learning: feature engineering, classification methods (random forest, Xgboost, SVM, Neural Networks),

clustering, topic modelling, text mining

Statistical Methods: time series, regression models, hypothesis testing, dimensionality reduction

Programming Languages: Python (sklearn, numpy, scipy, pandas), R, SQL, Shell Scripting, Java, Scala, Ruby

Software/Frameworks: Hadoop Ecosystem (Flume, Hive, MapReduce, Spark, Hbase, Sqoop, Solr), Kerberos, Linux,

Puppet, Docker, Oracle, MySQL, WebSphere, WebLogic, Cloudera, Hortonworks, IBM Big Insights