

# Nishant Arora

Data Scientist | BITS Pilani

Motivated Developer with experience in working with Data Science and Engineering as well as Software Development. Love to solve real life industry problems with data and code. Good at picking up new technologies quickly and getting things done.

## Contact

### Address

Panchkula, Haryana

### Phone

9929061890

### E-mail

nishant.arora.682@gmail.com

### WWW

www.nishantarora.ml

## Skills

Apache Spark	<div><div></div><div></div><div></div><div></div><div></div></div>
Python	<div><div></div><div></div><div></div><div></div><div></div></div>
Scala	<div><div></div><div></div><div></div><div></div><div></div></div>
Java	<div><div></div><div></div><div></div><div></div><div></div></div>
Machine Learning	<div><div></div><div></div><div></div><div></div><div></div></div>
SQL	<div><div></div><div></div><div></div><div></div><div></div></div>
Apache Airflow	<div><div></div><div></div><div></div><div></div><div></div></div>
MongoDB	<div><div></div><div></div><div></div><div></div><div></div></div>
Spark Streaming	<div><div></div><div></div><div></div><div></div><div></div></div>
AWS	<div><div></div><div></div><div></div><div></div><div></div></div>
Deep Learning	<div><div></div><div></div><div></div><div></div><div></div></div>
R	<div><div></div><div></div><div></div><div></div><div></div></div>
Elasticsearch	<div><div></div><div></div><div></div><div></div><div></div></div>
NLP	<div><div></div><div></div><div></div><div></div><div></div></div>
Git	<div><div></div><div></div><div></div><div></div><div></div></div>
Flask	<div><div></div><div></div><div></div><div></div><div></div></div>
CPP	<div><div></div><div></div><div></div><div></div><div></div></div>
A/B Testing	<div><div></div><div></div><div></div><div></div><div></div></div>
HTML/CSS/JS	<div><div></div><div></div><div></div><div></div><div></div></div>

## Software

IBM SPSS	<div><div></div><div></div><div></div><div></div><div></div></div>
MySQL	<div><div></div><div></div><div></div><div></div><div></div></div>
Workbench	<div><div></div><div></div><div></div><div></div><div></div></div>
PowerBI	<div><div></div><div></div><div></div><div></div><div></div></div>
Metabase	<div><div></div><div></div><div></div><div></div><div></div></div>
Autocad	<div><div></div><div></div><div></div><div></div><div></div></div>
Solidworks	<div><div></div><div></div><div></div><div></div><div></div></div>

## Work History

2020-01 -

2020-07

### Data Science Intern

*Ecom Express Pvt. Ltd., Gurgaon, Haryana*

- Worked on the implementation of an automated scalable end to end forecasting pipeline from fetching the raw data everyday to delivering the forecasts, using Apache Spark, Apache Airflow, Python, S3, MySQL.
- Worked on implementing pipelines which ingest raw as well as processed data from MySQL database to the S3 datalake which use various connectors between Apache Spark code, S3, MySQL and Python.
- Developed a module for Analysing Data by fitting various probability distributions to it using PySpark and Scipy.
- Worked on Locality extraction on a Pincode level by clustering on latitude and longitude values using Apache Spark and Scala.

2019-07 -

2019-12

### Data Science Intern

*couture.ai, Bangalore, Karnataka*

- Implemented a NLP Pipeline in Apache Spark and Scala which does various basic text cleanups to spell correction, part-of-speech based filtering, phrase extraction.
- Worked on the implementation of a pipeline which does automated feature selection from data.
- Implemented sales prediction pipelines using xgBoost and lightGBM models in Apache Spark and Scala.
- Worked on the implementation of recommendation and similar products pipeline using Topic Modeling in Apache Spark.
- Researched on AutoML and the implementation of such a platform.

2018-05 -

2018-07

### Software Development Intern

*nearbuy.com, Gurgaon, Haryana*

Developed an A/B Testing Framework based on Bayesian Statistics with a user friendly UI that shows test results in real time. Also devised a new Listing Ranking algorithm using Machine Learning which optimized various objectives important for the business.

2017-05 -

2017-07

### Software Development Intern

*nearbuy.com, Gurgaon, Haryana*

Worked on a Natural Language Processing Project : Designed a free text search parser using Apache Spark in Java. The classification model used was a Random Forest Classifier and the accuracy reached was 94%.

2016-05 -

2016-07

### Summer Intern

*H.M.T. Business Group, Panchula, Haryana*

## Languages

English

Hindi



Devised and Implemented an algorithm for the automatic generation of unique supplier code using C++ and MySQL which marked the end of the manual system which was in existence from years.

## Education

2014-08 -  
2020-07

**Master of Science - M.Sc. (Hons.) :  
Mathematics**

*Birla Institute Of Technology And Science , Pilani*

2014-08 -  
2020-07

**Bachelor of Engineering - B.E. (Hons.) :  
Mechanical Engineering**

*Birla Institute Of Technology And Science , Pilani*

## Accomplishments

**ACM ICPC Asia Amritapuri Site Regional Contest 2016**

**Participant** Team Name: treeBitsians

**Won the award for the Best Intern at nearbuy.com in the summer of 2018**

**Online Judges and Programming Contests:**

Hackerrank Algorithms Domain World Rank: 200, Percentile : 99%

Codeforces Rating: 1614, Problems Solved > 500

SPOJ World Rank: 2500

Secured 3rd position in CodeGenesis 1.0 on Hackerearth

Secured 4th position in College Hackathon on Hackerearth

Secured 8th position in Inter University Programming Contest 2016 at Plinth TechFest

Microsoft codefundo++ 2018 participant

**Personal Projects:**

- Implemented a visual search engine within elasticsearch which uses image vector extraction using the Resnet-18 Model, self implemented vector to string encoders and ranking algorithms.
- Developed a Flask Webapp which analyzes the person's intelligence levels through the input fingerprint ridge count and input values and generates a report of the same.
- Scraped advanced football statistics using BeautifulSoup from understat.com and used that to cluster the players based on their goal scoring and playmaking capabilities.
- Implemented various efficient 2D packing algorithms in MATLAB and Python.

**National Level Chess Player, Gold Medalist at state level**

## Certifications

2018-09

Deep Learning Specialization, Coursera

2020-04

Algorithms : Design and Analysis I and II, Stanford University

2017-11

Introduction to Machine Learning, Udacity

2019-01

Text Retrieval and Search Engines, Coursera

2018-06

SEO for Beginners, Udemy