



REMI SHAH

Big Data Engineer

ABOUT ME

Both in my working and personal life, I enjoy new experiences and challenges. The decision to take the opportunity to work for Norbulk Shipping aboard a Russian cargo ship for 7 weeks definitely had challenging consequences. Having never worked on a cargo ship previously nor having ever spoken Russian, I built close working relationships with my colleagues despite the language barrier and unfamiliarity with a cargo ship environment. This developed my ability to connect with people from different cultures and adapt to new working environments and resulted in me thoroughly enjoying my time on the ship

For the first half of 2016, after purchasing a drone and realising the extend of its filming capability, I started a business offering high definition, edited footage of venues, arenas and schools, filmed and edited for promotional purposes.

I obtained a CAA pilots licence to fly the drone and my colleague was responsible for video editing. To date videos for several schools, golf courses, public gardens and entertainment venues have been made.

CONTACT US

Kubrick Group

T: 020 3866 4620

E: consultants@kubrickgroup.com

W: kubrickgroup.com

OVERVIEW

I graduated from Cardiff University in Integrated Engineering in 2017. My third year dissertation was a study of methods used for real-time system identification of built structures such as buildings and bridges. It involved comparing methods for generating real-time health status reports and predictive fault reports for structures based upon current and historic data output by arrays of sensors mounted on the structures.

This was my first exposure to the concept and use of big data. The incredible accuracy of structural monitoring capabilities through using data was the catalyst for my passion in the field. I also studied the theory of Artificial Intelligence, notably genetic algorithms and neural networks, which is another exciting way for large data sets to be used – a great example is voice recognition.

WORK

DATA ENGINEER - KUBRICK GROUP - LONDON

September 2017 - Current

At Kubrick Group I gained knowledge and real world experience in professional skills, a wide yet in depth array of modern technical and analytic skills. We studied the agile software development lifecycles and gained detailed knowledge of platforms and infrastructure.

I specialised in Data Engineering which involved learning and developing advanced skills in Python, Spark, Hadoop, NoSQL and SQL as well as advanced Excel - these lead me to work towards accreditation as a Cloudera Hadoop CCA Developer.

As a junior data consultant I applied my skills on real client projects including:

- Seasonality correlation to sales for a large retail client.
- Feature analysis for a large financial services client providing HR with churn rate modelling against different employee profiles.

The projects were developed in agile teams using Git and Kanban boards following agile processes in the presence of an agile coach. Delivery of the projects used virtual environments and Docker. The professional skills I have acquired and matured were invaluable in these projects.

CONTROL ENGINEER – NORBULK SHIPPING LTD

August 2016 - September 2016

I was responsible for monitoring the condition of the engine room by manually checking machines and using computers outputting real-time sensor data. I then assessed if all pre-defined levels of efficiency and safety were met and if a change was required I was involved in the manual repair and replacement of the faulty components. The data analysed included flow rates, air pressures and temperature of fluids and gases within the engine network.

EDUCATION

CARDIFF UNIVERSITY: MENG INTEGRATED ENGINEERING – 1ST CLASS HONOURS

2013 - 2017

Overall degree mark: 72%

Two of my highest module marks were in Artificial Intelligence and Advanced Robotics where I achieved 91% and 85% respectively.

Final year mark: 78%.

For my final year project, I designed a ceramic 3D printer to create medical implants such as dental implants or bone scaffolds. In discussions with health professionals I worked with throughout the project, I was introduced to the idea of using big data in healthcare and the immensely beneficial insights it could generate for illness diagnosis.

KEY SKILLS

Python Libraries:

Pandas, Numpy, SK Learn, NLTK, PyMongo and Plotly

Cloudera Hadoop version 5.4:

HDFS, Yarn, Sqoop, Flume, Avro, Parquet, Impala, Hive and Hue

NoSQL:

MongoDB, Neo4J

SQL:

SSIS, Lavastorm, Advanced SQL

Apache Spark:

RDD, DataFrames, SparkSQL, PySpark with some exposure to and understanding of Scala

Source Control and IDE:

Git, SourceTree and Git Extensions, PyCharm

Other Skills:

Datalku, Linux CentOS 6, SSH, Putty, Agile