Neil John D. Ortega

neiljohn.ortega@gmail.com

8 Ligaya St., Brgy. Plainview, Mandaluyong City, 1550, Philippines

+63-977-185-9549

SKIILS AND COMPETENCIES

- **Soft Skills**: Strong work ethic, Good communication skills, Leadership and Teamwork skills, Collaborative, Highly adaptable, Problem solver, Critical thinker.
- **Machine Learning and Deep Learning**: Keras, Tensorflow, scikit-learn, nltk, pandas, numpy, etc.
- **Big Data**: Spark (including MLlib and Streaming), NiFi, Kafka, Storm, Hive, Hadoop, Cloudera distribution, etc.
- **Web Development and Visualization**: Django, django-rest-framework, Flask, Shiny, AngularJS, D3.js, etc.
- **DevOps**: Docker, Kubernetes, Jenkins, Ansible, AWS.
- **Scripting**: Python, R, SQL, Bash.
- **Collaboration and Project Management**: Git, JIRA, Confluence.

EXPERIENCE

- **Data Scientist**   Nov 2016 — Present
  Teradata Corporation

  - **Data Extraction from Scanned Documents**
    Pipeline for automatic extraction of desired field values from scanned (images and PDF files) version of documents. Involved in developing ad-hoc rule-based detection of some of the fields. Implemented image preprocessing scripts using OpenCV to improve image input to OCR service. Wrote scripts using Tesseract and Microsoft OCR to extract information from scanned images. Used ML techniques for predicting location of fields of interest in documents.

  - **Survey Response Classification**
    AI-based app for automated tagging of free-form customer survey responses with the goal of speeding up market defect discovery. Involved in developing the solution architecture. Wrote backend and frontend code for data ingestion and model results visualization using R, Python, and SQL. Trained and tuned LSTM + word-embedding models for the classification task using Keras with Tensorflow and scikit-learn. Created Docker-based pipeline for the team for efficient deployment to both local and production environments. Acted as domain expert for technologies the rest of the team are not familiar with.

  - **Customer Behavior Modeling**
    Hidden Markov model for predicting behavior of telecom subscribers for location-based and time-based targeting. Created parallelized Python pipeline to operationalize the model developed by data scientists. Coordinated with client's Cloudera Cluster administrators in tracking and debugging issues with their infrastructure. Benchmarked and optimized performance for both Hive data store and Teradata Aster compute engine that reduced end-to-end (E2E) processing time from several days to 2-3 hours.

  - **Model Monitoring and Maintenance**
    Created a Spark / Shiny application that allows users to run a scoring model against datasets in HDFS. Worked with the client's analytics team to properly scope the project and to determine parts of their workflow to be improved. Implemented dashboards for visualizing metrics indicating the health of the models. Deployed entire solution via Ansible and Cloudera Manager on the customer's AWS cluster. Solution streamlined the customer's processes and reduced E2E processing time from 1 week to 1 hour.

  - **Load Demand Forecasting and Disaggregation**
    Created an interactive Shiny application with a custom map UI via Leaflet that allows users to visualize both actual and forecasted electrical load demand and disaggregation. Presented a demo of the solution on-site.

  - **Real-Time Social Media Discovery**
    Real-time sentiment analysis application made using Apache NiFi, Kafka, Spark Streaming and SparkML (Machine Learning). Implemented both backend, using NodeJS and MongoDB, and frontend using Covalent/Angular and in-house chart widgets.

  - **Internal Consultant Directory**
    Custom internal directory for consultants connected to a recommender system meant for improving resource allocation and management. Wrote the REST API backend in Django and django-rest-framework.

  - **On-Site Search Improvements**
    Integrated pandas and numpy into Flask backend to improve its performance. Updated visualizations written with C3.js and D3.js.

  - **Plugin Development**
    Helped develop an internal reusable R plugin for creating customizable plots and widgets for Shiny applications.

- **On-boarding Materials**
  Created customized demo and exercise materials for Apache NiFi, Kafka, and Storm, integrating the three into a single streaming dataflow application in a Cloudera environment. Handled training of members from another department.

- **Senior Full-Stack Engineer**   Sep 2011 — Nov 2016
  icannhas Inc.

  - **Stock Market Buy-or-Sell Recommender**
    Machine learning-based approach for predicting stock market trend and recommending buy or sell orders. Implemented near real-time prediction backend using scikit-learn and Django. Implemented frontend using plot.ly and ReactJS.

  - **Consulting and (Internal) Product R&D**
    Worked on multiple web-based applications for both local and international clients. Implemented backend solutions mainly in Python, Django, django-celery, django-rest-framework, etc. Implemented frontend solutions using HTML5, CSS3, JS, and miscellaneous JS frameworks.

EDUCATION

- **Bachelor of Science in Civil Engineering, *magna cum laude***   2006 — 2011
  University of the Philippines, Diliman, Quezon City

ACTIVITIES AND ORGANIZATIONS

- **University of the Philippines Tennis Society**

  Secretary General for Special Events   2009 — 2010
  - Elected by the members as head of committee responsible in teaching tennis to future and current members and in handling the logistics of events.
  - Spearheaded annual university-wide tournament of the organization. Coordinated with sponsors, org members, and participants for the entire event.

  Secretary General for Campaigns and Propaganda   2008 — 2009
  - Elected by the members as head of committee responsible for promoting the organization and its activities and events.

- **University of the Philippines Programming Guild**

  Member   2009 — 2010
  10th Place 2009 ACM-ICPC Asia Manila Regional Contest