
HSBC

AI Cards Researcher

TIN-HSBC-001

About the Organisation

HSBC is one of the world's leading international banks. With a network that covers 64 countries and territories, HSBC offers you opportunities to develop your career in established and faster-growing markets. We serve our customers through our global businesses, which are supported by our operational and functional teams.

The Data and Architecture Office (DAO) was established to transform HSBC into a truly data-driven organisation by ensuring that every design decision has a built-in data consideration. In DAO, we will provide leadership that enables data and analytics to be a key driver of our business strategy. Working closely in partnership with senior stakeholders and delivery teams all over the world to connect the business processes, data, products, services, and technology that underpin customer value and financial success.

Group Business Intelligence & Analytics is a new Group-level function. Business Intelligence & Analytics unlocks additional value from our data providing insights to improve customer experience, increase revenue, optimise capital, enhance risk management, deliver regulatory compliance and reduce costs.

Group Business Intelligence & Analytics will offer the following services:

- End-to-end production of management information, reporting and business intelligence.
- End-to-end production of business analytics and data integrity analytics use cases.
- A self-service visualisation platform.
- A self-service analytics & artificial intelligence platform.

Role Description and Responsibilities

The Intern's role is to support further roll out of AI Cards version 1, and the development of AI Cards version 2. Specifically:

- Be involved in the technical aspects of responsible and ethical AI such as explainability, transparency, bias, and unintended AI decision making;
- Be responsible for supporting work on the target operating model and will interact with the stakeholders to map out existing and design future user processes;
- Be responsible for engaging with technical colleagues to collaborate and guide the new system build;
- Developing assets and thought leadership for responsible and ethical AI.

Turing Internship Network – Spring 2022

In partnership with the Global Businesses and Functions you could be involved in activities such as:

- Solving challenging business problems using advanced machine learning methods such as Deep Learning and quantitative analytics;
- Cleaning, aggregating and interpreting data in preparation for analysis;
- Partnering with developers and engineers to assess and provide assurance of AI deployed to deliver business value;
- Defining approaches to embed and scale ML models;
- Building reusable assets, solutions and developing best practices for current and future business problems;
- Consulting on complex analyses and advanced ML methods

Expected Outcomes

As HSBC embraces technology as part of its strategic plan to “Digitise at scale”, we continue to see an increased usage of data analytics, Artificial Intelligence (AI) and Machine Learning (ML) across the Bank. External to HSBC, there have been significant developments in recent years with government agencies and regulators around the world making clear that new laws will soon shape how companies use AI. Additionally, in order to operationalise the HSBC Principles for Ethical Use of Big Data and AI, and to implement the developing AI Framework, HSBC must know what AI has been deployed and where. Accordingly, there is increasing need for the Bank to support appropriate governance, monitoring, and risk management of our AI products and services.

The above points emphasise the requirement for HSBC to have a more complete register or “inventory” of AI products and services.

Project Outcome: AI Cards Inventory

An “AI Card” is a structured document which accompanies an AI use case to record associated key facts. Information captured includes the context in which the AI is intended to be used, details of performance, data sources, and potential ethical concerns including metrics that capture bias, fairness and inclusion considerations.

Using AI Cards helps provide increased transparency around AI tools that benefits all, regardless of expertise. Engineers may use AI Cards to understand whether AI could be reused or fine-tuned for their needs and to ensure they are aware of weaknesses and biases. Business colleagues can use AI Cards as a reference to help explain the use of data and AI tools to clients. This is essential information which HSBC does not otherwise currently capture in a consistent and standardised manner.

The implementation of AI Cards, and mandating their use across the Bank, is helping the management of AI and associated risks.

Functional requirements created have allowed for the development of an initial system hosted on the HSBC Google Cloud Platform that is now being rolled out globally across all Businesses and Functions to catalogue the existing AI assets within HSBC.

For the Intern, their outcomes will be:

- Deep practical knowledge of AI implemented and deployed in financial services – taking the theory you will have learned in academia to how it is applied in industry to achieve business value.

Turing Internship Network – Spring 2022

- Proficiency in at least one high level programming language - Python [e.g. Tensorflow].
- Experience in developing techniques to support algorithm auditing via initial research to practice of assessing, mitigating, and assuring an algorithm's safety, legality, and ethics. This area encompasses current research in areas such as AI Fairness, Explainability, Robustness, Privacy, as well as matured topics of Data ethics, management and stewardship.

Supervision and Mentorship

The Intern will work in an agile scrum team alongside Cloud Engineers, Architects and Data Scientists to deliver support further roll out of AI Cards version 1, and the development of AI Cards version 2.

The Intern will be supervised by an experienced Innovation Research Lead, they will be part of the BI&A team. The team will provide them with a network of people to consult with as well as provide exposure to different technologies and use cases.

Ideal Intern

We are specialists at collecting and analysing large data sets, and we use this knowledge to help teams across the Bank spot and understand both opportunities and challenges. You will use your investigative and problem-solving skills to analyse data and uncover trends that deliver actionable insights for the Business.

This role is suited for people who have a technical and commercial mind-set and are excellent communicators. A successful candidate is a problem solver who enjoys exploring data, is excited by difficult modelling challenges, and possesses strong communication skills to effectively interface with the multi-disciplinary teams.

We are looking for dedicated and curious people who are:

Data enthusiasts. This programme will develop your statistical knowledge, passion for data or any programming skills you may already hold. You will gain knowledge of the theoretical grounds of artificial intelligence, machine learning and mathematical frameworks.

Committed researchers. Our markets and client needs are constantly changing. You should always be curious; questioning ideas and showing a keen interest in learning about our methods.

Agile workers. You will work in a fast-paced environment where everything evolves very quickly. You will maintain momentum by promoting agile working and ensuring we constantly deliver for our clients.

Analytical thinkers. You will need a good eye for detail, a creative mind for analysing trends, and the courage to adopt new analytical methods.

Team players. We collaborate by listening to new ideas, as well as voicing our own, in order to deliver and develop the best service for our clients.

Internship Logistics

Salary: £30,000 p.a pro rata.

Internship length: 6 months, full time.

Location: Remote with occasional travel to London to work alongside colleagues or for training.

HSBC

AI Researcher

TIN-HSBC-002

About the Organisation

HSBC is one of the world's leading international banks. With a network that covers 64 countries and territories, HSBC offers you opportunities to develop your career in established and faster-growing markets. We serve our customers through our global businesses, which are supported by our operational and functional teams.

The Data and Architecture Office (DAO) was established to transform HSBC into a truly data-driven organisation by ensuring that every design decision has a built-in data consideration. In DAO, we will provide leadership that enables data and analytics to be a key driver of our business strategy. Working closely in partnership with senior stakeholders and delivery teams all over the world to connect the business processes, data, products, services, and technology that underpin customer value and financial success.

Group Business Intelligence & Analytics is a new Group-level function. Business Intelligence & Analytics unlocks additional value from our data providing insights to improve customer experience, increase revenue, optimise capital, enhance risk management, deliver regulatory compliance and reduce costs.

Group Business Intelligence & Analytics will offer the following services:

- End-to-end production of management information, reporting and business intelligence.
- End-to-end production of business analytics and data integrity analytics use cases.
- A self-service visualisation platform.
- A self-service analytics & artificial intelligence platform.

Role Description and Responsibilities

The Intern's role is to support the Global AI Enhancement Programme. Specifically:

- Be involved in the technical aspects of responsible and ethical AI such as explainability, transparency, bias, and unintended AI decision making;
- Be responsible for supporting work on the target operating model and will interact with the stakeholders to map out existing and design future user processes;
- The candidate will be responsible for engaging with technical colleagues to collaborate and guide the new system build;
- Developing assets and thought leadership for responsible and ethical AI.

In partnership with the Global Businesses and Functions you could be involved in

Turing Internship Network – Spring 2022

activities such as:

- Consulting on complex analyses and advanced ML methods
- Partnering with developers and engineers to assess and provide assurance of AI deployed to deliver business value
- Defining approaches to embed and scale ML models

Expected Outcomes

As HSBC embraces technology as part of its strategic plan to “Digitise at scale”, we continue to see an increased usage of data analytics, Artificial Intelligence (AI) and Machine Learning (ML) across the Bank. Much of this AI is developed internally for business benefit, and also features increasingly heavily in third party (external vendor) services and tools that HSBC uses.

Given the Bank’s increasing use of AI, there is a need to understand the scale and nature of AI across HSBC, and to proactively manage AI risks. As financial institutions adopt AI at scale there is increasing concern and regulatory focus on the effectiveness of AI governance to avoid adverse reputational, financial, legal and ethical outcomes, and to enable safe adoption. At HSBC, AI risk is a composite risk that is broader than just Model risk. AI risk also impacts compliance, technology, legal, operational, data and other risks.

HSBC has already made some significant steps in AI risk management with areas showcasing thought leadership and best practice elements. However, this AI management has grown in a grassroots fashion without coordination across the Bank and has resulted in a fragmented governance landscape.

This new project is a new formal **Global AI Enhancement Programme** set up to deliver a horizontal risk management approach that brings together stakeholders and framework elements from our different areas.

Specifically, this project will enhance AI management including and not limited to the following:

1. Operating Model & Implementation: enhancement of elements developed in silos that have to date led to an inconsistent approach across HSBC
2. Roles and responsibilities: formalise risk ownership from current state of not fully established and executed as “side of desk” work
3. Definition of AI: create and agree definitions that can be applied consistently across the Bank whilst ensuring compliance with evolving regulations
4. AI-related risk appetite: develop guidance on the level of risk HSBC is willing to take through the adoption of AI
5. Policies and procedures including third party risk and vendor management to remove inconsistencies in identification and treatment of third party AI
6. Development & testing including data: guidance around data quality, privacy and non-traditional data
7. Ongoing assessments & validation: develop practical guidance on ongoing monitoring of risks associated with the use of AI

For the Intern, their outcomes will be:

- Deep practical knowledge of AI implemented and deployed in financial services – taking the theory you will have learned in academia to how it is applied in industry to

Turing Internship Network – Spring 2022

achieve business value.

- Proficiency in at least one high level programming language - Python [e.g. Tensorflow].
- Experience in developing techniques to support algorithm auditing via initial research to practice of assessing, mitigating, and assuring an algorithm's safety, legality, and ethics. This area encompasses current research in areas such as AI Fairness, Explainability, Robustness, Privacy, as well as matured topics of Data ethics, management and stewardship.
- Deep knowledge on AI risk management at a complex global organisation.

Supervision and Mentorship

The Intern will work in an agile scrum team alongside Cloud Engineers, Architects and Data Scientists to deliver the global enhancement AI management programme.

The Intern will be supervised by an experienced Innovation Research Lead, they will be part of the BI&A team. The team will provide them with a network of people to consult with as well as provide exposure to different technologies and use cases.

Ideal Intern

We are specialists at collecting and analysing large data sets, and we use this knowledge to help teams across the Bank spot and understand both opportunities and challenges. You will use your investigative and problem-solving skills to analyse data and uncover trends that deliver actionable insights for the Business. This role is suited for people who have a technical and commercial mind-set and are excellent communicators. A successful candidate is a problem solver who enjoys exploring data, is excited by difficult modelling challenges, and possesses strong communication skills to effectively interface with the multi-disciplinary teams.

We are looking for dedicated and curious people who are:

Data enthusiasts. This programme will develop your statistical knowledge, passion for data or any programming skills you may already hold. You will gain knowledge of the theoretical grounds of artificial intelligence, machine learning and mathematical frameworks.

Committed researchers. Our markets and client needs are constantly changing. You should always be curious; questioning ideas and showing a keen interest in learning about our methods.

Agile workers. You will work in a fast-paced environment where everything evolves very quickly. You will maintain momentum by promoting agile working and ensuring we constantly deliver for our clients.

Analytical thinkers. You will need a good eye for detail, a creative mind for analysing trends, and the courage to adopt new analytical methods.

Team players. We collaborate by listening to new ideas, as well as voicing our own, in order to deliver and develop the best service for our clients.

Internship Logistics

Salary: £30,000 p.a pro rata.

Internship length: 6 months, full time.

Location: Remote with occasional travel to London to work alongside colleagues or for training.

HSBC

Synthetic Data Engineer

TIN-HSBC-003

About the Organisation

HSBC is one of the world's leading international banks. With a network that covers 64 countries and territories, HSBC offers you opportunities to develop your career in established and faster-growing markets. We serve our customers through our global businesses, which are supported by our operational and functional teams.

The Data and Architecture Office (DAO) was established to transform HSBC into a truly data-driven organisation by ensuring that every design decision has a built-in data consideration. As DAO we will provide leadership that enables data and analytics to be a key driver of our business strategy. Working closely in partnership with senior stakeholders and delivery teams all over the world to connect the business processes, data, products, services, and technology that underpin customer value and financial success.

Group Business Intelligence & Analytics is a new Group-level function. Business Intelligence & Analytics unlocks additional value from our data providing insights to improve customer experience, increase revenue, optimise capital, enhance risk management, deliver regulatory compliance and reduce costs.

Group Business Intelligence & Analytics will offer the following services:

- End-to-end production of management information, reporting and business intelligence.
- End-to-end production of business analytics and data integrity analytics use cases.
- A self-service visualisation platform.
- A self-service analytics & artificial intelligence platform.

Role Description and Responsibilities

The Synthetic Data Engineer Intern is responsible for developing GCP PaaS (Platform as a Service) for synthetic data generation, solving cutting edge problems and working with the rest of the team to productionise the solution. This would be done using the already existing blocks and API, however generalisation and scaling to enterprise level will require creativity. We are looking for people who are interested in cloud engineering (namely GCP), and have curiosity for innovation and AI/ML.

The Synthetic Data Engineer Intern will partner with architecture and cloud teams to develop practical solutions and custom modules for Synthetic Data as a Service and will need to ensure that all cloud solutions follow security and compliance controls.

The Synthetic Data Engineer Intern will work in a fast-paced environment where everything

Turing Internship Network – Spring 2022

evolves very quickly and will maintain momentum by promoting agile working and ensuring we constantly deliver for our clients. We collaborate by listening to new ideas, as well as voicing our own, in order to deliver and develop the best service for our clients.

Expected Outcomes

Synthetic data refers to artificial information created to serve as substitute data for developers and businesses.

[Gartner](#) predicts that by 2024, 60% of the data used for the creation of AI and analytics projects will be synthetically generated.

To respond to the growing needs, HSBC initiated **Synthetic Data as a Service** project. Its goal is to create capability for synthetic datasets generation, that preserve both the utility and privacy of our data. The generated synthetic data would be used for rapid prototyping and experimentation internally and support external collaboration with start-ups, academics and technology partners.

The approach to achieve this to combine internal and external capability, develop in-house synthetic engine and expertise, and combined with leading academia and vendors.

Currently, we have several synthetic data generation tools for structured and unstructured data, conceptual architecture with data flows and approvals, and API for Synthetic Data as a Service user (to upload real data, get information on the capabilities, and report on privacy and utility evaluation). We also collaborate with the Alan Turing Institute on several synthetic tools creation.

The outcome being to pilot Synthetic Data as a Service PaaS in HSBC this year.

Supervision and Mentorship

The synthetic data engineer intern will work in an agile scrum team alongside Cloud Engineers, Architects and Data Scientists to deliver GCP PaaS (Platform as a Service) for synthetic data generation.

The Intern will be supervised by an experienced Innovation Research Lead, they will be part of the Data Science Chapter. The Chapter will provide them with a network of people to consult with as well as provide exposure to different technologies and use cases.

Key skills to be gained from the role:

- Real-world, hands-on experience using Google Cloud technologies (cloud infrastructure, API, web services) essential for future career as Cloud Engineer
- Experience in platform development in a large financial enterprise
- Experience in synthetic data generation using advanced Machine Learning methods (GANs, autoencoders)

Ideal Intern

The ideal Intern will have a passion for data to improve customer experiences and change businesses.

- They will have a foundational level knowledge of cloud platforms and data management and security;
- basic/intermediate coding ability in a number of computer languages (e.g. Python);

Turing Internship Network – Spring 2022

- strong problem-solving, written and oral communication skills.

Internship Logistics

Salary: £30,000 p.a pro rata.

Internship length: 6 months, full time.

Location: Remote with occasional travel to London to work alongside colleagues or for training.

HSBC NLP Engineer TIN-HSBC-004

About the Organisation

HSBC is one of the world's leading international banks. With a network that covers 64 countries and territories, HSBC offers you opportunities to develop your career in established and faster-growing markets. We serve our customers through our global businesses, which are supported by our operational and functional teams.

The **Data and Architecture Office (DAO)** was established to transform HSBC into a truly data-driven organisation by ensuring that every design decision has a built-in data consideration. As DAO we will provide leadership that enables data and analytics to be a key driver of our business strategy. Working closely in partnership with senior stakeholders and delivery teams all over the world to connect the business processes, data, products, services, and technology that underpin customer value and financial success.

Data Insights and Incubation is a new Group-level function. Our vision is to equip our customers with knowledge and understanding through appropriate use of data. We aim to drive decisions and actions from the very top of the organisation downwards by providing leaders with information and tools to do the right thing.

Focused on NLP, this role will provide an opportunity to work in a collaborative agile environment, benefit from a network of world-class experts, and leverage industry-leading cloud and data technologies.

Role Description and Responsibilities

The Intern will be responsible for building models to extract insights from unstructured text data and identify themes in large data sets using NLP to interpret the text component. You will support the wider team in designing and productionising a process to simplify the generation of insights from multiple data sets. This may be done using existing NLP tools such as Google USE, however application and scaling to enterprise level will require creativity.

We are looking for candidates who are interested in cloud engineering (namely GCP), and have curiosity for innovation and AI/ML. The Intern will need to ensure that all cloud solutions follow security and compliance controls.

Expected Outcomes

On a daily basis HSBC generates a colossal amount of text information. Often this information requires human interpretation to drive a decision or an action, but at the scale that the data is generated, automation becomes increasingly important in optimising decision making. Many of the data sets that are used by the team to generate insights include a text

Turing Internship Network – Spring 2022

component, for example the description of services provided by a vendor as part of an enterprise contract.

The aim of the project is to create a capability to highlight patterns in these data sets based on the interpretation of text and provide recommendations to executives where a tangible benefit may be realised. An example use case would be highlighting where multiple suppliers are contracted to provide similar services and where the organisation may benefit from consolidation.

The project will be delivered via a combination internal and external capabilities, with the objective of building out in-house knowledge, supported by leading academia and vendors.

Supervision and Mentorship

The Intern will work in an agile team alongside Data Engineers, Analysts, ML Engineers and other Data Scientists to deliver the end-to-end use case.

The Intern will be supervised by an experienced Data Scientist specialised in NLP. The Insights team will provide a support network of domain experts to consult with as well as provide exposure to different technologies and use cases.

Ideal Intern

The ideal candidate will have a passion for data science and using data and ML to deliver tangible value for customers.

- They will have a strong statistical background with practical experience in coding languages, e.g., Python (preferred), R, Scala.
- Knowledge of machine learning modelling techniques and how to fine-tune those models
- Experience using specialized machine learning libraries, e.g. Tensorflow, PyTorch, Fast.ai
- Familiarity with NLP techniques (NER, semantic search, sentence encoding)
- Ability to demonstrate how they have understood and applied new techniques in the field of machine learning as they emerge
- Strong verbal/written communication and data presentation skills.

This role will be working on GCP and building models with unstructured data, so prior experience with cloud technologies and in building models with these data sets is advantageous.

Internship Logistics

Salary: £30,000 p.a. pro rata.

Internship length: 6 months, full time.

Location: Remote with occasional travel to either London, Birmingham or Sheffield to work alongside colleagues or for training.

HSBC

Sustainability Data Researcher

TIN-HSBC-005

About the Organisation

HSBC is one of the world's leading international banks. With a network that covers 64 countries and territories, HSBC offers you opportunities to develop your career in established and faster-growing markets. We serve our customers through our global businesses, which are supported by our operational and functional teams.

Wholesale Banking: Our Wholesale Big Data Lake is the largest aggregation of data ever within HSBC. We have over 300 sources which equate to more than 20PTB of data, with a use case portfolio of over 110 projects that span all the business lines within Wholesale. We are utilising the latest machine learning tools and technologies to solve these hypotheses and deliver value and truly unique insights.

Wholesale Data Analytics, Data Science and Engineering: The team in the Wholesale Chief Data and Analytics Office is working to deliver an ecosystem of curated, enriched and protected sets of data – created from global, raw, structured and unstructured sources. The Intern selected will be working for wholesale analytics, specifically in the Research and Development pillar under the Wholesale Chief Data Analytics Office (WCDAO). The placement will be within the sustainability pod under the R&D pillar and working specifically for the Wholesale Environmental, Social and Governance as a service pod (WESGaaS). The sustainability pod is a cross functional team of product owners, data scientists, analysts and engineers as well as business analysts, business owners from various functions of the bank. The pod works in collaboration with various business owners as well as Group in HSBC.

Our Data Analytics Ecosystem solves complex problems using cutting-edge technologies; helping to rapidly implement insights from data that can help drive more informed decision making. We deploy smart machines to process complex and large sets of data, impossible in the legacy manual mining methods. Data underpins everything we do; from risk & regulatory management, through monetisation, to predicting client behaviour.

Role Description and Responsibilities

The Intern will have the option of working on one of two projects within the Sustainability Pod.

Project 1: Extend the scoring model for entities within the HSBC universe as well as unbanked entities to determine sustainability scores. This is going to be a composite model that involves quantitative data such as internal HSBC data, external ESG vendor data, geospatial analysis, remote sensing with satellite imaging as well as qualitative data. The qualitative data will need research to transform into usable features. The sustainability score will be derived based on a weighted version of the models. It will also help provide more data driven estimates of sector, geography based ESG scores to help with the big data gap problem in ESG. The overall sustainability scores will be used extensively in various applications across the bank such as

Turing Internship Network – Spring 2022

climate risk, sustainable supply chains, etc.

OR

Project 2: This project will work on the natural language modelling suite for sentiment analysis on vendor and frontline engagement/comments, real time news, social media content analysis along with analysing sustainability reports to help build a multi classification model for the impact to UN sustainability goals. The project will aim to build the library with various NLP models including transformers such as BERT for sustainability which will contribute towards HSBC's IP in natural language processing.

Expected Outcomes

The sustainability performance of a company is increasingly as important to its key stakeholders as its financial metrics.

To achieve this, part of the problem that requires solving is the ability to successfully measure, monitor, and control emissions across the value chain. In Wholesale Data and Analytics, we are harnessing the power of Data and AI to contribute to our HSBC Environmental, Social, Governance (ESG) commitments and targets around this.

One such initiative, partnering with the Data and Architecture Office, is the Wholesale Environmental, Social, Governance as a Service (WESGaaS).

WESGaaS has Relationship Managers and customers at the heart. Various external ESG market data points were collected to combine with insights/research, industry precedence and transaction history to sit alongside existing client insights to create a HSBC ESG dashboard that provides an overall view of our Global Banking (GB) and Commercial Banking (CMB) clients sustainability scores and metrics. Through machine learning we've been working on estimating the emissions and other data points to solve the problem of coverage. Some of the methods we've commenced exploring include using cutting edge techniques like geospatial analysis and sentiment analysis via AI research in the fields of deep learning and natural language processing (NLP). Simply put, accurately gather data covering the full scope of emissions, to create a baseline from which they can determine a customer's ESG performance. Through this smart use of our data, we can target opportunities to support their transition to a low carbon model and identify commercial opportunities for our clients. ESG-as-a-Service will initially focus on Global Banking clients, before cascading the industry and geography level insights into clients where appropriate.

Intern Outcomes

For both projects, Interns will complete their placement with:

- Sustainable finance knowledge
- Experience with machine learning models
- Working in an experienced Research & Development team
- Experience of theoretical and academia application across various businesses
- Valuable knowledge in ESG data and latest estimation models for GHG emissions

Supervision and Mentorship

The Intern will work in an agile team alongside Data Engineers, Analysts, ML Engineers and other Data Scientists to deliver either Project 1 or 2.

The Intern will be supervised by an experienced Research Lead in Sustainability and Data. The Sustainability Pod will provide a support network of domain experts to consult with as

Turing Internship Network – Spring 2022

well as provide exposure to different technologies and use cases.

Ideal Intern

The ideal Intern will have a lot of motivation and energy to learn new things and keep up with the ever-evolving space of sustainability. Specific skills and interests to include:

- Strong statistical background with practical experience in coding languages, e.g., Python and sql skills
- Machine Learning fundamentals (deep learning knowledge within ML is a plus for NLP based use cases)
- Team player
- Strong data visualisation / presentation skills
- Strong verbal and written communication skills
- An interest in climate change/sustainability

Internship Logistics

Salary: £30,000 p.a. pro rata.

Internship length: 6 months, full time.

Location: Remote with occasional travel to either London, Birmingham or Sheffield to work alongside colleagues or for training.

HSBC

Data Scientist

TIN-HSBC-006

About the Organisation

HSBC is one of the world's leading international banks. With a network that covers 64 countries and territories, HSBC offers you opportunities to develop your career in established and faster-growing markets. We serve our customers through our global businesses, which are supported by our operational and functional teams.

HSBC UK is transforming from a data-rich to data-driven organisation; the Data and Analytics department are central to this transformation. The department delivers data analytics, machine learning (ML) and management information which are used to measure, automate and improve the performance of the business. The department is at the centre of customer intelligence and marketing; using ML to communicate the right message, to the right customer, at the right time.

The team is leading the adoption of cloud and open-source technologies using ML in innovative ways to automate processes, delivering efficiencies in banking operations and improving customer engagement.

Role Description and Responsibilities

The Data Scientist Intern is responsible for exploring and finding meaning in large and complex data sets. Across our brands we have 14 million customers generating petabytes of structured and unstructured data. The Data Scientist Intern will work in our data and analytics environment in Google Cloud Platform (GCP). They will use open source (Python and Spark) and Google native Artificial Intelligence (AI) engines to explore these data sets, develop and deploy machine learning models that change how we interact with our customers and how we run our banking operations.

The Data Scientist Intern will be conversant in and will use techniques such as natural language processing, neural network techniques to analyse documents and speech data, use reinforcement learning and other deep learning techniques.

The Data Scientist Intern will work in an agile scrum team alongside Data Engineers, Analysts, ML Engineers and other Data Scientists to deliver end-to-end use cases.

Expected Outcomes

The Data Scientist Internship will be focused on developing ML models to automate a manual sales quality checking process. This process currently involves agents manually reviewing documentation, listening to calls and analysing structured data sources, to assess whether a product sale was the right outcome for the customer.

Turing Internship Network – Spring 2022

The project aims to develop ML techniques to replace the manual assessment of the documentation and speech data to automatically check and validate the sale.

The business outcomes will be to reduce the time for the quality checking process from weeks to minutes, to extend the coverage of the quality checks from a sample of 10% today to 100% of sales; and most importantly to improve the outcomes for customers, ensuring the right product is sold to every customer.

Supervision and Mentorship

The Data Scientist Intern will work in an agile scrum team alongside Data Engineers, Analysts, ML Engineers and other Data Scientists to deliver the end-to-end use case.

The Intern will be supervised by an experienced Data Scientist, they will be part of the Data Science Chapter. The Chapter will provide them with a network of people to consult with as well as provide exposure to different technologies and use cases.

Ideal Intern

The ideal Intern will have a passion for data science and using data and ML to improve customer experiences and change businesses.

- They will have a strong statistical background with practical experience in coding languages, e.g., Python (preferred), R, Scala.
- Knowledge of machine learning modelling techniques and how to fine-tune those models, e.g., Deep Learning / Neural Networks, especially in relation to speech and text extraction from documents.
- Experience using specialized machine learning libraries, e.g., Fastai, Keras, Tensorflow, PyTorch, Sci-kit Learn.
- Demonstrate how they have understood and applied new techniques in the field of machine learning as they emerge.
- Strong verbal/written communication and data presentation skills.

This role will be working on GCP and building models with unstructured data, such as speech and documentation, so prior experience with cloud technologies and in building models with these data sets is advantageous.

Internship Logistics

Salary: £30,000 p.a. pro rata.

Internship length: 6 months, full time.

Location: Remote with occasional travel to either London, Birmingham or Sheffield to work alongside colleagues or for training.