The National Archives

The National Archives use Microsoft Cognitive Services and Databricks to identify and select digital assets to be stored

With ever growing amounts of born-digital information reaching the age of selection or deletion, there are not enough well-trained eyes available to review the required documents. The risk of losing records of great historical importance becomes more and more prevalent. The National Archives have therefore identified a critical need to establish machine learning techniques as common practice for record management in many Government departments.

The results of this research conclude that the architecture deployed by Adatis can locate and classify documents of historical importance in a way that is cost efficient, easy to operate by non-technical users and flexible to a wide spread of document types.

Situation.

Legislation states that after 20 years government departments need to determine if the asset should be stored or not. With vast volumes of 'born digital records' The National Archives and Government departments could no longer manually identify assets to be stored.

They wanted to advise other government departments on how technology could assist them in identifying and selecting relevant assets more quickly and so they did not miss anything.

Approach.

The National Archives decided to work with Adatis in order to conduct the research and build of a proof of concept into machine learning techniques, to classify digital assets that should be selected for preservation.

Adatis used an ensemble of Microsoft Cognitive services and custom ML models, trained using Azure Databricks, to search the documents for key phrases, entities and core topics. All of which were used to infer the need for selection.

Outcome.

The research paper and proof of concept developed using Azure Machine Learning technologies will enable The National Archives to advise Government departments on ways to efficiently identify and select digital assets and documents that need to be stored. It will also ensure that, in line with legislation, critical assets are not missed and reduce the time, overhead costs and manual effort requires to identify these assets.

Adatis offer professional services specialising in data analytics, from data management strategy and consultancy through to world class delivery and managed services.

With offices In London, Farnham and Sofia they have been delivering innovative and successful solutions for clients since 2006.

Data Science

Data as a Service

Modern Data Warehouse

Data Architecture

Data Analytics

Data Strategy

We love helping people unlock the power and value of their data.

If you are interested in finding out more please contact us on **+44 (0)1252 267 777** or via email at **enquiries@adatis.co.uk**

You can also check out our website at **www.adatis.co.uk** or come and see us at numerous global conferences and speaking events.

adatis