Bloomberg

DATA AUTOMATION

University of Maryland — QUEST HONORS PROGRAM, Quality Enhancement Systems and Teams

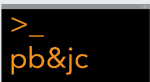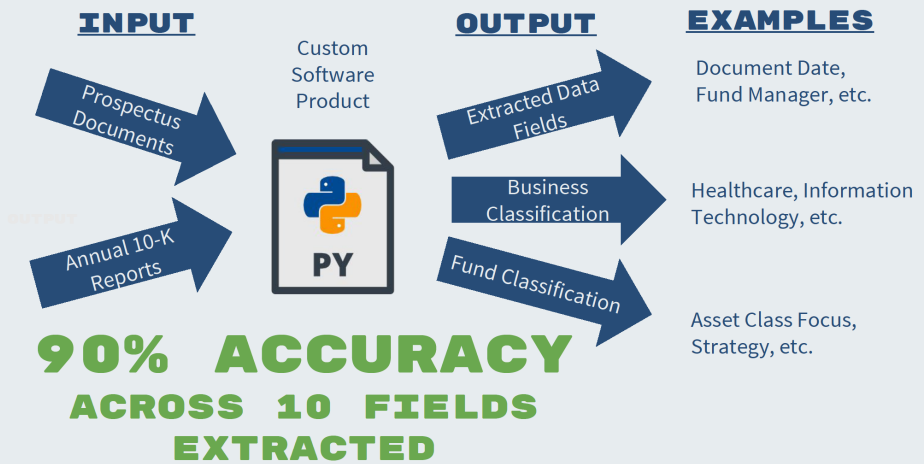We've been given the opportunity to provide **value** to Bloomberg by **automating:**
- **Extraction** of financial data from prospectus documents
- **Classification** of businesses and funds

Our software enables Bloomberg to deliver **accurate** financial data to its customers.

1000 FUNDS
PDF

PER WEEK

X

5 MINUTES
FIELD EXTRACTION

=

83 ANALYST HOURS SAVED
PER WEEK

Our software product frees analysts for other tasks by solving several challenges:
- PDF documents do not store structured data, making text extraction unreliable.
- Each fund prospectus is unique, making uniform automation difficult.
- The English language has subtleties that are difficult to teach a machine.

INPUT
Prospectus Documents

Annual 10-K Reports

Custom Software Product

PY

OUTPUT
Extracted Data Fields

Business Classification

Fund Classification

EXAMPLES
Document Date, Fund Manager, etc.

Healthcare, Information Technology, etc.

Asset Class Focus, Strategy, etc.

90% ACCURACY
ACROSS 10 FIELDS EXTRACTED

>_
pb&jc

Pete Dziki
Mechanical Engineering

Bobby Ellett
Finance & Marketing

Jeremy Krach
Computer Science

Christian Johnson
Computer Science

STUDENT TEAM: PB&J CONSULTING (PBJC)

Pete Dziki Mechanical Engineering	Bobby Ellett Finance & Marketing	Christian Johnson Computer Science	Jeremy Krach Computer Science
Project Champion: Cristiane Coutinho, Funds Product Manager			
Faculty Advisor: Dr. Jim Purtilo			

PROJECT SUMMARY

Bloomberg L.P. is a financial software, data, and media company headquartered in New York, NY, that conducts business around the globe. Through the Bloomberg Terminal financial software system, Bloomberg connects financial professionals with the real time market data that allows them to make educated investment decisions. Bloomberg seeks to increase workflow efficiency by automating the manual process of both data extraction and entity classification from financial documents. Bloomberg aims to utilize machine learning and software-based solutions to maximize efficiency in this automation process and has enlisted PBJC to design and deliver a software data extraction tool. Through this tool, Bloomberg will speed up data delivery to customers and free up analyst time that can be allocated toward other tasks.

CONTRIBUTIONS AND RECOMMENDATIONS

The software package delivered to Bloomberg employs a combination of business rules and natural language algorithms to create a powerful data extraction tool. PBJC identified business rules through financial research, using prospectus financial documents and 10-K annual reports to detect patterns in phrasing and document layout. These rules served as the basis for the software development phase of the project. The team developed a comprehensive software package that extracts key data fields and information relevant to classification from financial documents. The development team designed innovative algorithms that make use of Natural Language Processing (NLP) and machine learning to extract data and perform classification with high accuracy. The algorithms were uniquely tuned for Bloomberg's use case and account for sentence structure and common keywords to produce a scoring system to rank and extract relevant data. Testing determined that the software package delivered to Bloomberg extracts data at a 90.35% accuracy rate. This is estimated to free a total of 83 work hours per week. The team is currently testing classification accuracy to deliver even greater value.