Shoe Style & Size Detection System

UWC Robert Sobukwe Rd, Bellville Cape Town, 7535

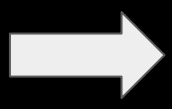
Supervisor: M. M. SUNGAY

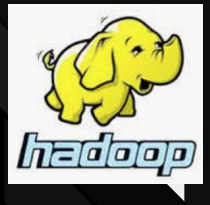
Member:

Wanga Mbabe
Jinwei Liu
Lwando Maciti
Heinrich Davids

By Gamma Group

OVERVIEW





Data Analysis

Intelligent Supply Chain

Better Decision Making

Jinwei

Literature (Project feasibility)

- Feasibility Financial → No cost required to develop the system
- Technical Feasibility → Google colab , Airtable , python libraries and Draw.io
- Risk Feasibility → Customer related risks
- Resource and Time Feasibility →
  - Programming device (Laptop) ,
  - Programming tools (freely available)
  - Time to complete all task

Literature

Big Data Engineering Technologies

- **HDFS** - used by hadoop for reliable storage and access to data
- **MapReduce -** used for parallel processing of large amounts of data running in many clusters.
  - Map Function: divides grouped data into multiple splits and breaks elements int key value tuples.
  - Reduce Function: retrieves the input from the map () function and joins the data tuples into smaller tuples
- **Sqoop** – coordinate the transfer of data between Hadoop and a database system in a parallel manner
- **Spark** – distributed computing engine for parallel processing and analysis of a

(*Batch Processing*)

(*Stream Processing*)

Lwando

Literature

UNIVERSITY of the WESTERN CAPE

Similar system/project

- Solving shoe classification and retrieval problem by using CNN.
- Integrated supply chain management based on CoT. Comparison to the traditional supply chain.
- Hadoop (MapReduce, filesystem, statistical analysis) and Big Data
- Apache Spark(Low latency, fast processing speed, real-time) vs Hadoop(High latency, Mapreduce, batch processing)

Jinwei

Literature

UNIVERSITY of the WESTERN CAPE

60 YEARS of hope, action & knowledge

BD Security and Privacy

BD Security

- Internal equipment and processes
- Culture, Strategies, Structure
- External data sources

BD Privacy

- Capacity to restrict and control
- Data being used responsible
- Control which data are being collected
- Understanding my rights w.r.t data collected

Privacy Policy

DATA COLLECTION FOR MARKETING PRIVACY POLICY

Last updated [July, 15, 2020]

INTRODUCTION

SHOE CITY ("we" or "us" or "our") respects the privacy of our users ("user" or "you"). This Privacy Policy explains how we collect, use, disclose, and safeguard your information when you visit our stores. Please read this privacy policy carefully. If you do not agree with the terms of this privacy policy, please do not access the site.

We reserve the right to make changes to this Privacy Policy at any time and for any reason. We will alert you about any changes by updating the "Last Updated" date of this Privacy Policy. Any changes or modifications will be effective immediately upon posting the updated Privacy Policy on the Site, and you waive the right to receive specific notice of each such change or modification.

You are encouraged to periodically review this Privacy Policy to stay informed of updates. You will be deemed to have been made aware of, will be subject to, and will be deemed to have accepted the changes in any revised Privacy Policy by your continued use of the Site after the date such revised Privacy Policy is posted.

COLLECTION OF YOUR INFORMATION

We may collect information about you when you enter our stores for marketing purposes only.

Personal Data

Non-Personally identifiable information will be collected ads you enter our stores. The information collected will only be of your shoes/ feet, and from this we cannot identify any personal information about you.

USE OF YOUR INFORMATION

Having accurate information about you permits us to provide you with a smooth, efficient, and customized experience. Specifically, we may use information collected about as you enter our stores:

- Compile anonymous statistical data and analysis for use internally only
- Increase the efficiency and operation in our stores
- Monitor and analyze usage and trends to improve your experience when you visit our stores.
- Respond to product and customer service requests.

DISCLOSURE OF YOUR INFORMATION

We may share information we have collected about you in certain situations. Your information may be disclosed as follows:

By Law or to Protect Rights

If we believe the release of information about you is necessary to respond to legal process, to investigate or remedy potential violations of our policies, or to protect the rights, property, and safety of others, we may share your information as permitted or required by any applicable law, rule, or regulation. This includes exchanging information with other entities for fraud protection and credit risk reduction.

Heinrich

Project Plan

The project is divided into three phases.

- Phase 1: Preliminary and data collection.
- Phase 2: ML model development and Hadoop configuration.
- Phase 3: Integration of ML and Hadoop.

Jinwei

Problem Statement

UNIVERSITY of the WESTERN CAPE

- Historic data from POS causes delay of information.
- Lack of information about potential customers.

Jinwei

Experimental Setup

UNIVERSITY of the WESTERN CAPE

60 YEARS of hope, action & knowledge

The system aims to improve the efficiency of the supply chain and timely reports about each individual store.

Camera	+	ML model	+	Hadoop	=	Shoe System
Data capturing Sensor		Classification Obtain useful info		Data storage Data analysis		Intelligent Supply Chain

Jinwei

Classification Model

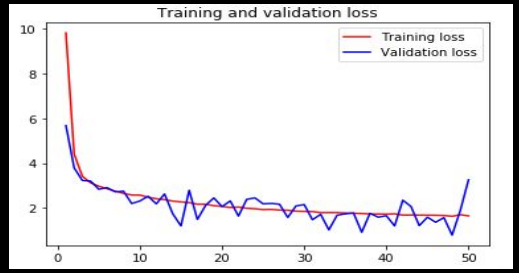
UNIVERSITY of the WESTERN CAPE

50 YEARS of hope, action & knowledge

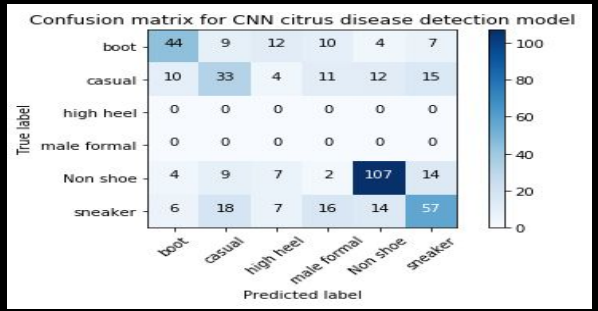
2D CNN Model Summary

```
Model: "sequential_3"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_11 (Dense)             (None, 150, 150, 16)      64
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 150, 150, 32)      4640
_________________________________________________________________
max_pooling2d_5 (MaxPooling2 (None, 75, 75, 32)        0
_________________________________________________________________
dropout_9 (Dropout)          (None, 75, 75, 32)        0
_________________________________________________________________
conv2d_6 (Conv2D)            (None, 75, 75, 64)        18496
_________________________________________________________________
max_pooling2d_6 (MaxPooling2 (None, 37, 37, 64)        0
_________________________________________________________________
flatten_3 (Flatten)          (None, 87616)             0
_________________________________________________________________
dropout_10 (Dropout)         (None, 87616)             0
_________________________________________________________________
dense_12 (Dense)             (None, 256)               22429952
_________________________________________________________________
dropout_11 (Dropout)         (None, 256)               0
_________________________________________________________________
dense_13 (Dense)             (None, 128)               32896
_________________________________________________________________
dropout_12 (Dropout)         (None, 128)               0
_________________________________________________________________
dense_14 (Dense)             (None, 64)                8256
_________________________________________________________________
dense_15 (Dense)             (None, 6)                 390
=================================================================
Total params: 22,494,694
Trainable params: 22,494,694
Non-trainable params: 0
_________________________________________________________________
```

Training and los curve

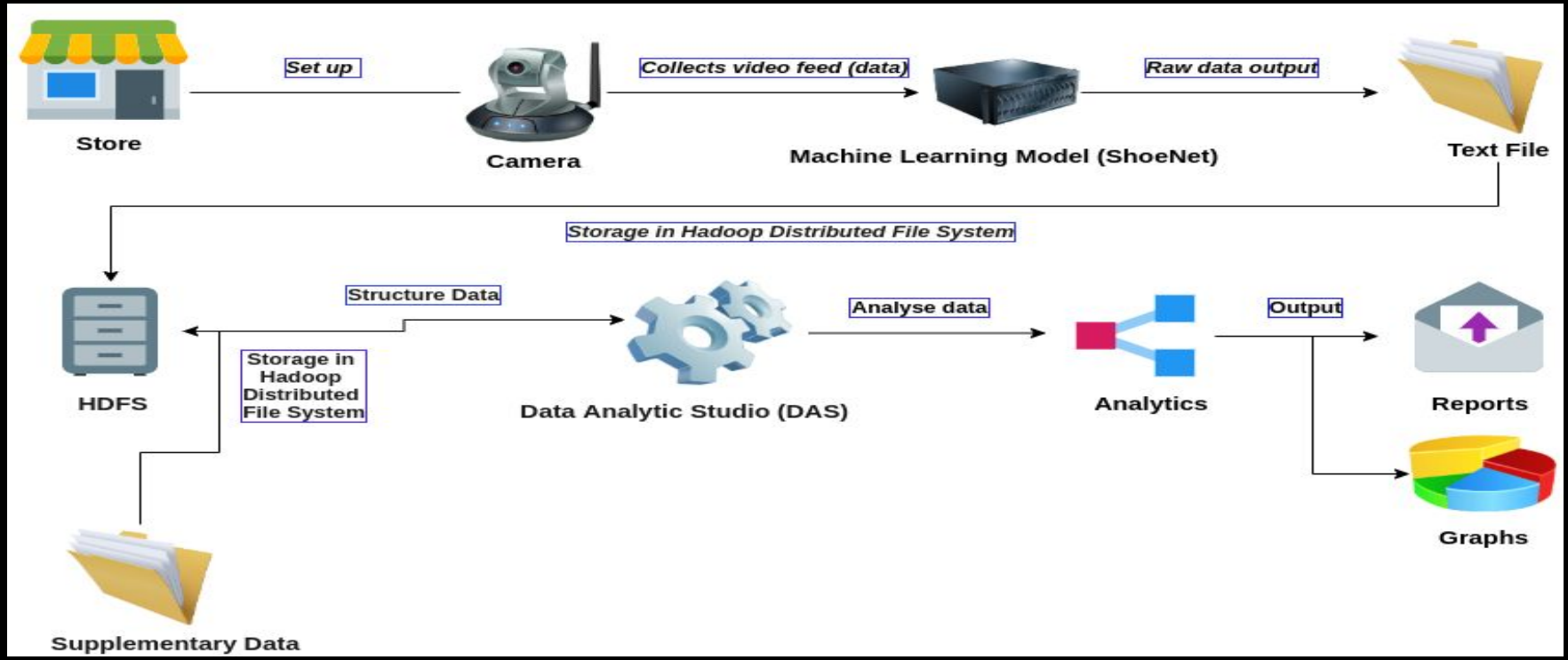


Confusion Matrix

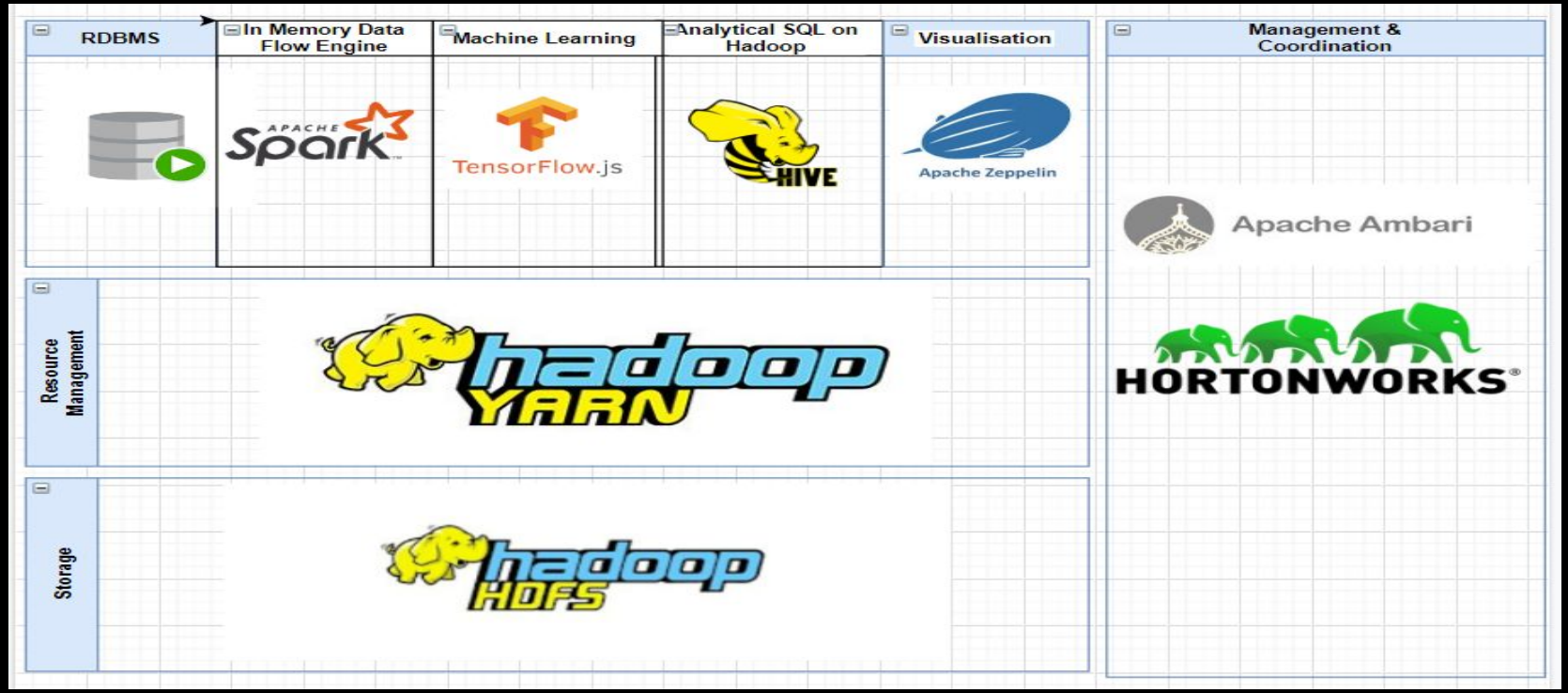


Lwando

Proposed Solution Diagram



Proposed Solution Architecture



Business Value and Process

- Traditional POS data vs Big Data
- No Sale = No information
- What does POS data tell us about our customers?
- Can BD give more insight?
- Does BD replace or add value?

Business Value and Process

UNIVERSITY of the WESTERN CAPE

60 YEARS of hope, action & knowledge

Data ML Model Collected

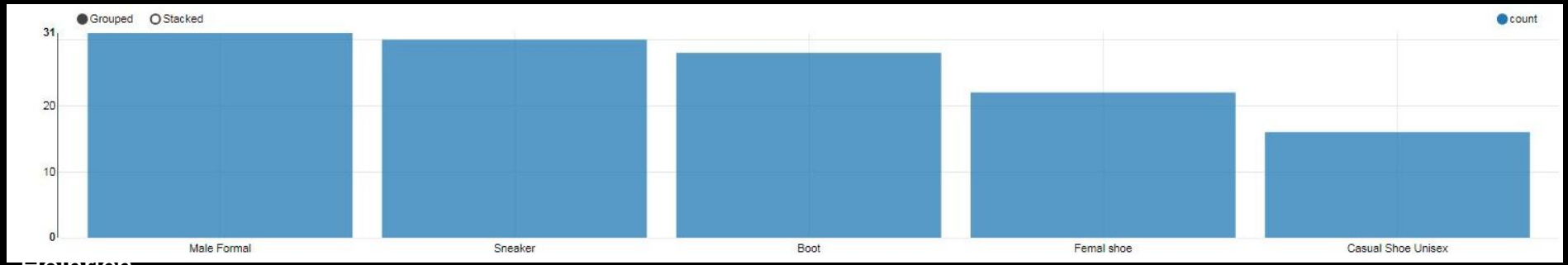
```
%spark2.sql

SELECT case when word = 'male_formal' then 'Male Formal'
            when word = 'sneaker' then 'Sneaker'
            when word = 'boot' then 'Boot'
            when word = 'high_heel' then 'Femal shoe'
            when word = 'casual' then 'Casual Shoe Unisex' end as shoe_type, count(*) as count FROM words GROUP BY word having count(*) > 1  ORDER BY count DESC
```

SPARK JOB FINISHED

settings

shoe_type	count
Male Formal	31
Sneaker	30
Boot	28
Femal shoe	22
Casual Shoe Unisex	16



Heinrich

Business Value and Process

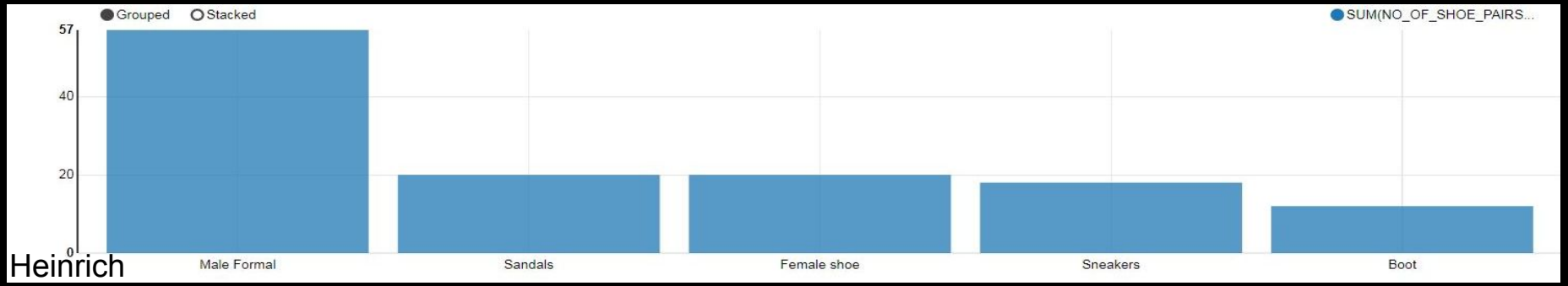
PO/ Actual data

```
%spark2.sql
-- Display word counts in descending order
SELECT *  FROM postylecounts ORDER BY 2 DESC
```

SPARK JOB FINISHED

settings

STYLE_DESC	SUM(NO_OF_SHOE_PAIRS)
Male Formal	57
Sandals	20
Female shoe	20
Sneakers	18
Boot	12



Heinrich

Tools and Technologies

UNIVERSITY of the WESTERN CAPE

60 YEARS of hope, action & knowledge

Software

- Hadoop
- Google Colab
- Zeppelin
- Spark
- Hortonworks
- HIVE
- OpenCV
- Jupyter Notebook
- Python
- Tensorflow API
- CSS
- HTML
- Virtual Machine

Hardware

- Laptop
- USB
- IP Camera
- Mobile router
- UPS
- Smartphone

Lwando

Future Development

UNIVERSITY of the WESTERN CAPE

60 YEARS of hope, action & knowledge

- Build better CNN model to detect and capture shoe preciously.
- The distance between the shoe and the camera may cause the measurement less accurate.
- Change the capturing technique from CNN model to weight mat.
- Facial recognition can be used to differentiate the gender of shoe owner.

Jinwei

References

[1] Wei Fang, Xue Zhi Wen, Yu Zheng, and Ming Zhou. A Survey of Big Data Security and Privacy Preserving. IETE Technical Review (Institution of Electronics and Telecommunication Engineers, India), 34(5):544–560, 2017.

[2] Emmanuel Sirimal Silva, Hossein Hassani, Dag Øivind Madsen, and Liz Gee. Googling fashion: forecasting fashion consumer behaviour using google trends. Social Sciences, 8(4):111, 2019.

[3] Technological, Organizational and Environmental Security and Privacy Issues of Big Data: A Literature Review. Procedia Computer Science, 100:19– 28, 2016.

[4] Mike Ananny. Toward an ethics of algorithms: Convening, observation,probability, and timeliness.Science, Technology, & Human Values, 41(1):93–117, 2016.

[5] K. Eze, “The Essence of Data Engineering,” no. November, pp. 2–4, 2018.

[6] S. Shahrivari, “Beyond batch processing: Towards real-time and streaming big data,” Computers, vol. 3, no. 4, pp. 117–129, 2014, DOI: 10.3390/computers3040117.

[7] Singh K and Kaur R 2014 Hadoop: Addressing challenges of Big Data *Souvenir 2014 IEEE Int. Adv. Comput. Conf. IACC 2014* 686–9

[8] Khosla N and Venkataraman V Building Image-Based Shoe Search Using Convolutional Neural Networks

[9] Jaggi H S and Kadam S S 2016 Integration of Spark framework in Supply Chain Management *Procedia Comput. Sci.* **79**

[10] D. Namiot, “On Big Data Stream Processing,” Int. J. Open Inf. Technol., vol. 3, no. 8, pp. 48–51, 2015, [Online]. Available: http://injoit.org/index.php/j1/article/view/225.

[11] Awad, Hussain & Nassar, Mohammad. (2010). Supply Chain Integration: Definition and Challenges. Lecture Notes in Engineering and Computer Science. 2180.

[12] Khosla N and Venkataraman V Building Image-Based Shoe Search Using Convolutional Neural Networks

[13]