## **Mathematics & Technology**

A Brief of Data Scientist







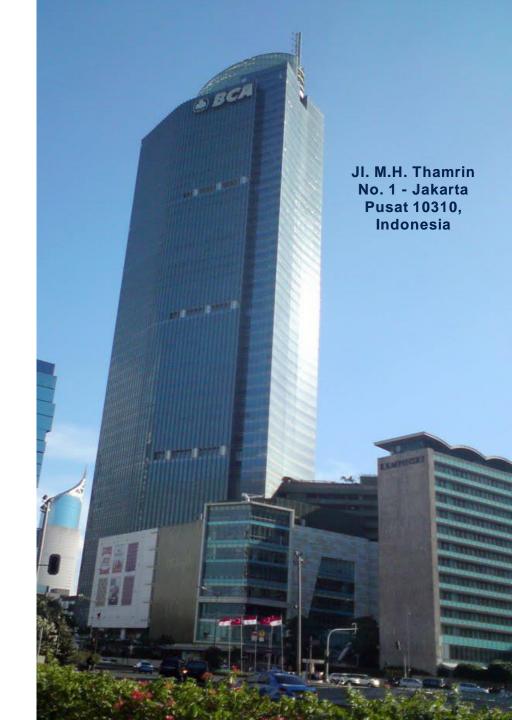
Happy Christanti, S.Si.

2016 : Graduated (Bachelor Degree of Mathematics at Sanata Dharma University)

2016 : Work at Areti (Education Company)

2017: Work at BCA (Strategic Information Technology Group / Data Management)

Phone (62-21) 23588000 ext. 16321 Email <a href="mailto:happy\_christanti@bca.co.id">happy\_christanti@bca.co.id</a>





### **Banking Industry**













### **Major Needed**





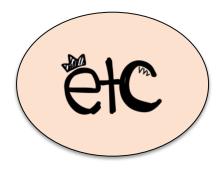














#### Data

Costumers



Apps





ATM

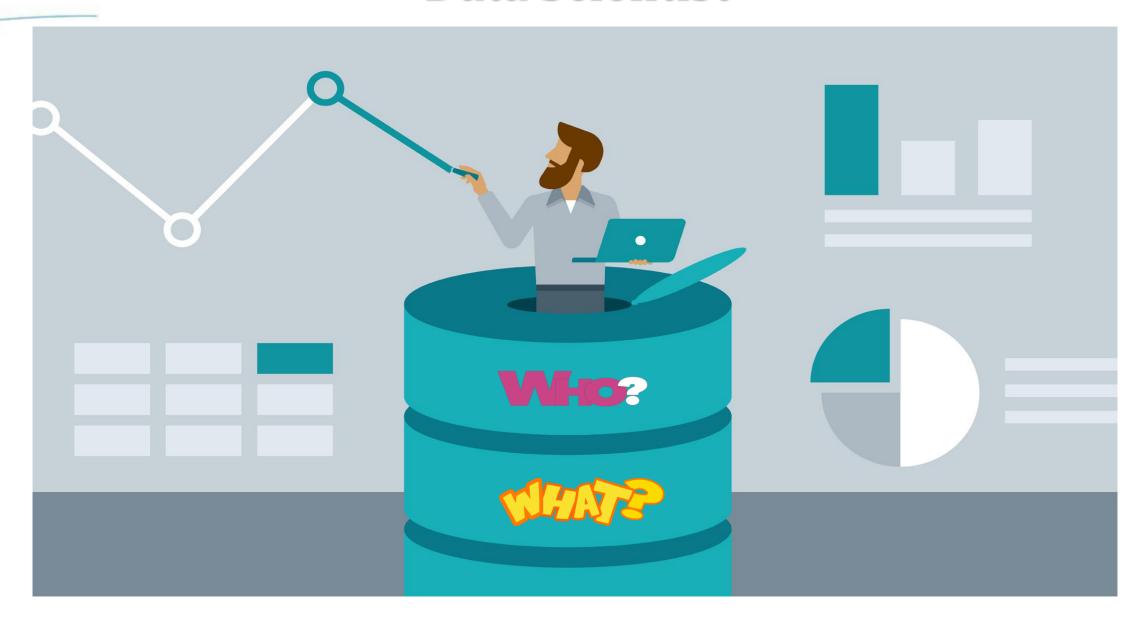


Merchant

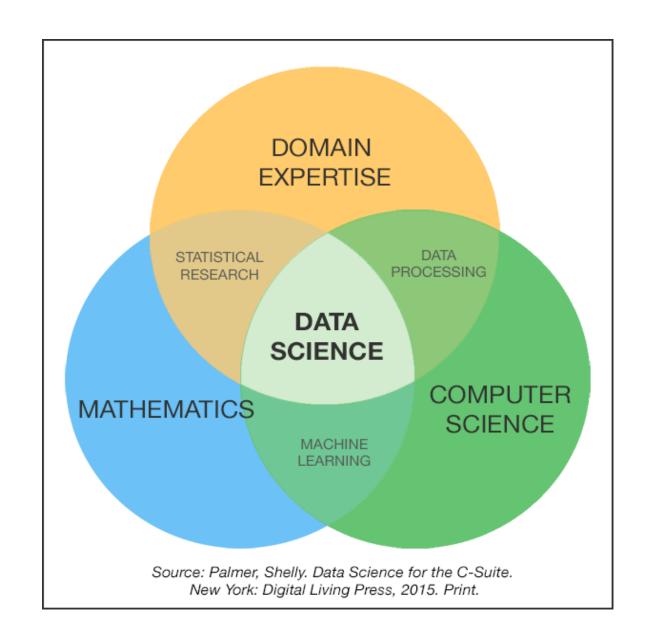


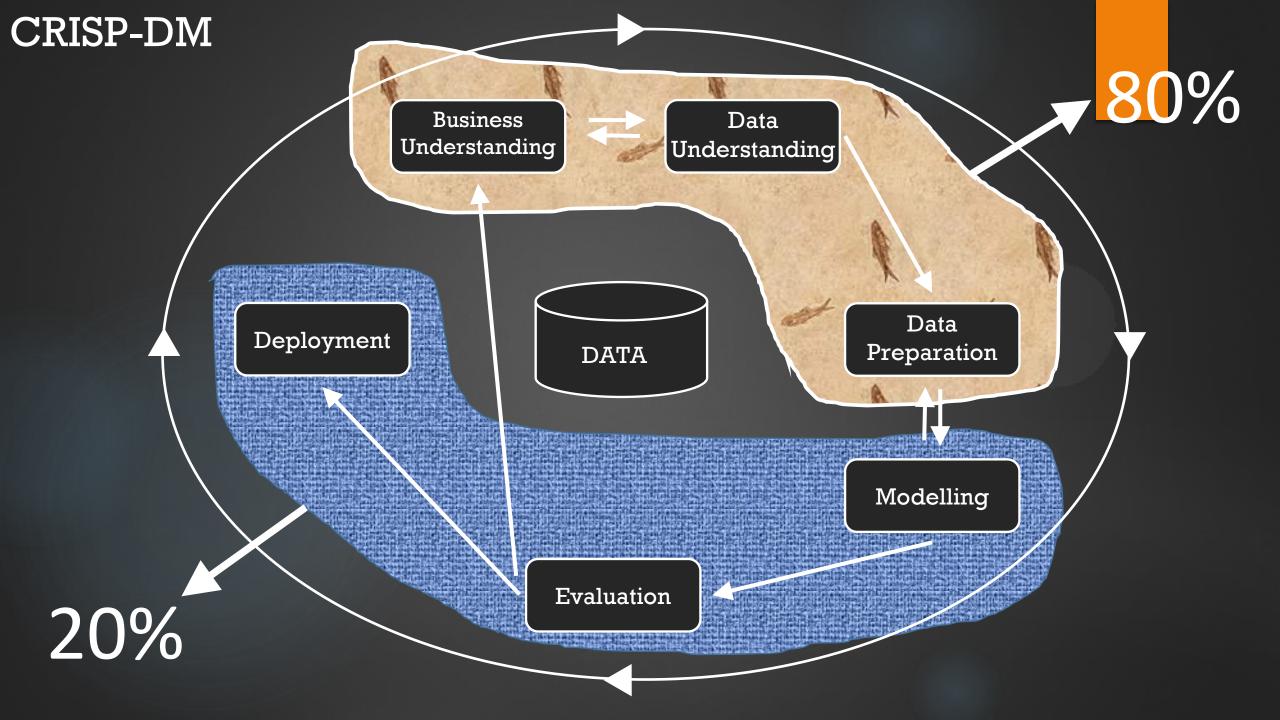


## **Data Scientist**





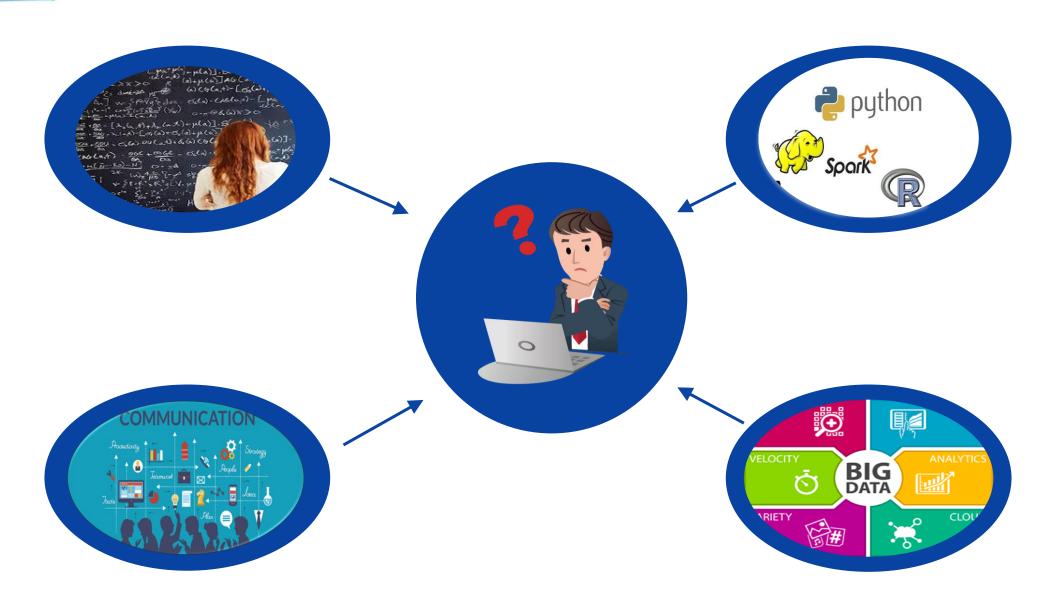




#### Data Science Skillset 9. Data Munging Sampling Using ETL Stratified How much Data? Principal Component Analysis Google OpenRefine 5. Text Mining / NLP Data Surve 8. Data Ingestion Clustering Using Mahout 6. Visualization Using Perceptron Data Exploration in R (Hist, Boxplot etc) Linear Regression Uni, Bi & Multivariate Viz Regression Ranking ggplot2 Logistic Regression Histogram & Pie (Uni) Tree & Tree Map 10. Toolbox Author: Swamilassification drasekaran 4. Machine Learning Scatter Plot (Bi) MS Excel w/ Analysis ToolPak Line Charts (Bi) Prob Den Fn (PDF) 1. Fundamentals Java, Python Spatial Charts ANOVA R, R-Studio, Rattle Survey Plot Skewness Matrices & Linear Algebra Fundamentals Weka, Knime, RapidMiner Timeline **Continuos Distributions** Hash Functions, Binary Tree, O(n) (Normal, Poisson, Gaussian) ecision Tree Hadoop Dist of Choice Lists Data Frames Cumul Dist Fn (CDF) Spark, Storm Relational Algebra, DB Basics **Factors** Reading CSV Data Random Variables Flume, Scibe, Chukwa Inner, Outer, Cross, Theta Join Arrays Reading Raw Data Bayes Theorem CAP Theorem Nutch, Talend, Scraperwiki Matrices ( Subsetting Data Probability Theory Tabular Data Zookeeper Webscraper, Flume, Sqoop Data Frames & Series Vectors ( Manipulate Data Frames Avro Percentiles & Outliers Rapid Name & Data Nodes Storm: Hadoop ( ) tm, RWeka, NLTK Sharding ( Histograms ( Functions Setup Hadoop (IBM / Cloudera / HortonWorks) Variables ( RHIPE Rhadoop, OLAP ( IBM SPSS **Exploratory Data Analysis** RHIPE Data Replication Principles D3.js, ggplot2, Shiny Expressions Multidimensional Data Model Descriptive Statistics rmr mean, median, range, SD, Var) R Basics Cassandra ETL **Hadoop Components** R Setup MongoDB, Neo4i Map Reduce Fundamentals R Studio Reporting Vs BI Vs Analytics 7. Big Data 2. Statistics 3. Programming



#### **Skills of Data Scientist**





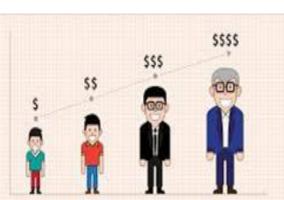


#### **Predictive Analytics in Banking Industry**



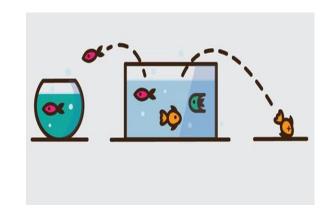
**Fraud Detection** 







**Collections Scoring** 





Marketing Optimization



Churn Analysis

Market Basket Analysis



#### Online Courses for Data Scientist



# DataCamp COURSEICE





# I skate to where the puck is going to be, not where it has been. Wayne Gretzky

# Thank you

Any question?

