

Data Science for DBA – Get Started Now

Manohar Punna

Data Platform Consultant, Data Platform MVP

LobsterPot Solutions

Vice President – DataPlatformGeeks



Manohar Punna

- Blogger, Speaker, Mentor –Community Enthusiast and Data Platform MVP.
- Microsoft Data Platform Consultant at LobsterPot Solutions.
- Authored **One DMV a Day** Series –Wrote 80 blogs in 80 days.
<http://bit.ly/OneDMVaDay>
- Website: www.ManoharPunna.com
- Speaker at **SQLBits, SSGAS, SQL Server Day, Data Platform Day, SQL Saturdays.**
- **Vice President** at DataPlatformGeeks
- **@ManuSQLGeek**

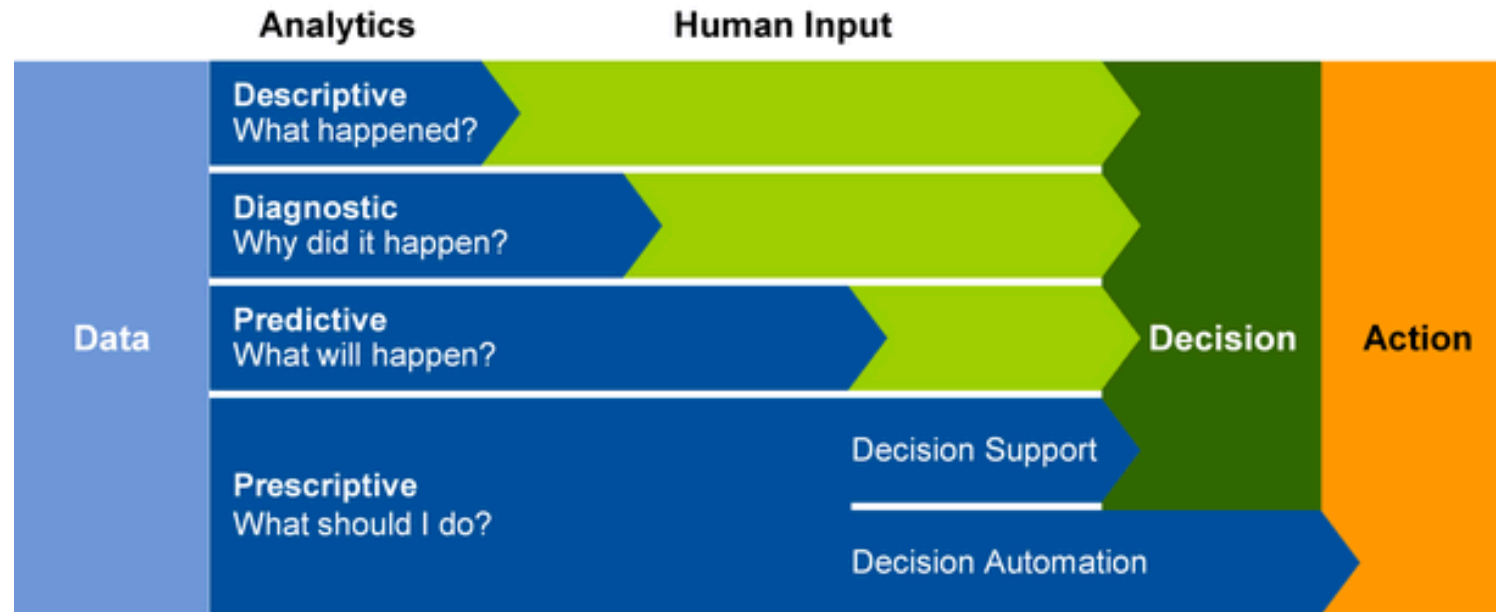


Agenda

- Advanced Analytics
- Data Science Team Process
- Roles – Machine Learning Engineer, Data Engineer
- R in SQL Server
- Using R Studio
- Azure Machine Learning Studio



Advanced Analytics



"Data scientists are not business analysts. They are professionals with the capability to derive mathematical models from data to reap clear and hard-hitting business benefits. They need to network well across different business units and work at the intersection of business goals, constraints, processes, available data and analytical possibilities."

[Source: Gartner \(October 2014\)](#)



Data Science Team Process

- Business Understanding
 - Define Objectives and identify data sources
- Data Discovery and Acquisition
 - Ingest, explore and update data
- Modeling
 - Creating and training models
- Deployment
 - Operationalize
- Business Acceptance
 - Testing and validation/ re-train and re-score



Machine Learning Engineer

- Python & R programmers
- Build Models
- Statistics
- Neural Networks
- Deep Learning
- Machine Learning Algorithms



Data Engineer

- Advanced DBA/BI Dev
- Both structured and unstructured data
- Less modeling and more data stuff
- Data Wrangling –
 - Discovery, cleansing and preparing data
 - Can be used for model input



Starting with R/Python

- Language for statistics computing and graphics
- Built at Statistics Department at University of Auckland
- Data manipulation, calculation and graphical display
- Microsoft introduced R into SQL Server with 2016 version
- Python is now supported in SQL Server 2017



Demo - 1

Using R in SQL Server



R Studio

- R Studio is an IDE for R
- Open source supporting the open source cause to deliver R
- Supports direct code execution, plotting, history, debugging and workspace management



Demo - 2

Using R Studio



Starting with Azure ML Studio

- Azure ML Studio is a Microsoft cloud hosted service
- Helps you build, deploy and share predictive analytics solutions
- Easy to learn and use models available
- Integrated with R and Python modules



New to machine learning?

If you're just getting started, read David Chappell's [Introduction for technical professionals](#).

Get a quick introduction to data science in [five short videos](#) and learn how to select the best algorithm with the [ML cheat sheet mini site](#).



New to the product?

If you have data science skills, but are getting started with Azure Machine Learning, take a look at our [tutorials](#), or get started with sample experiments from the [Gallery](#). More questions? Ask our engineers in our [forum](#).



Current R or Python user?

Azure Machine Learning Studio includes hundreds of built-in packages and support for custom code.

Learn more about how to get started with Machine Learning with R and Python by reading our [blog](#).



Data scientist or developer?

Azure Machine Learning is designed for applied machine learning. Use best-in-class algorithms and a simple drag-and-drop interface, and go from idea to deployment in a matter of clicks. [Try it for free](#).

If you're a developer who wants the data science to be built-in, take a look at our [APIs](#) and Azure Marketplace.

[Source: Azure Machine Learning Studio](#)



Demo - 3

Overview of Azure ML Studio

Creating and Deploying model in Azure ML Studio

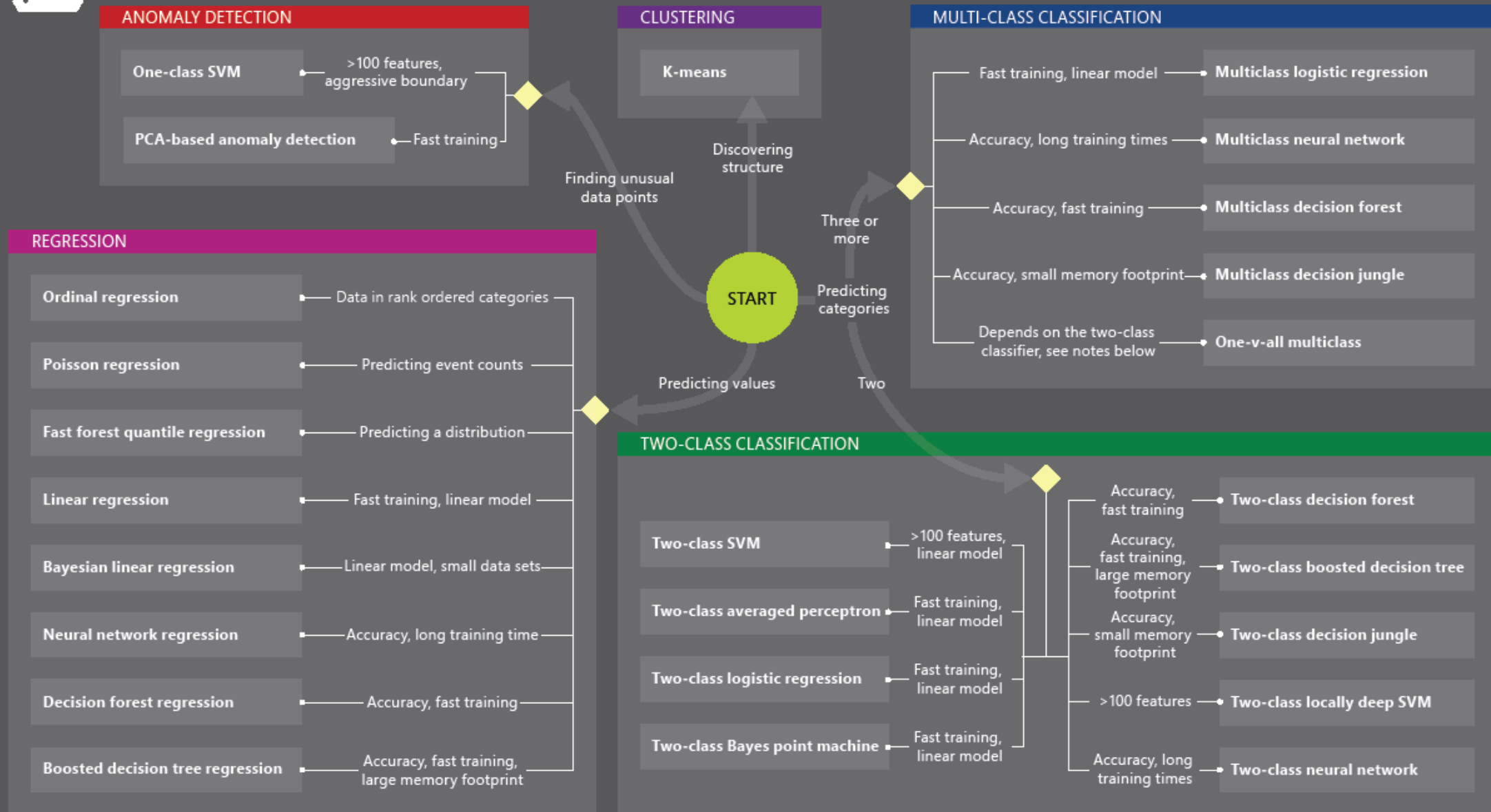
Consuming the service and visualize in Power BI





Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.



Summary

- Explore R and Python inside SQL Server
- Choose the right tools to suit the requirement
- R Studio, Power BI, Azure ML Studio
- Explore Azure ML using different algorithms

References

- Azure ML and Data Science git repo - <https://github.com/Azure/Azure-MachineLearning-DataScience>
- Next Steps: Azure ML Workbench - <https://docs.microsoft.com/en-us/azure/machine-learning/preview/quickstart-installation>
- Data Science VM on Azure Portal



Just like Jimi Hendrix ...

We love to get feedback

Please complete the session feedback forms



SQLBits - It's all about the community...

Please visit Community Corner, we are trying this year to get more people to learn about the SQL Community, equally if you would be happy to visit the community corner we'd really appreciate it.

