Team Artifical Triangles
E-commerce – SmartKart

Recommendation Engine

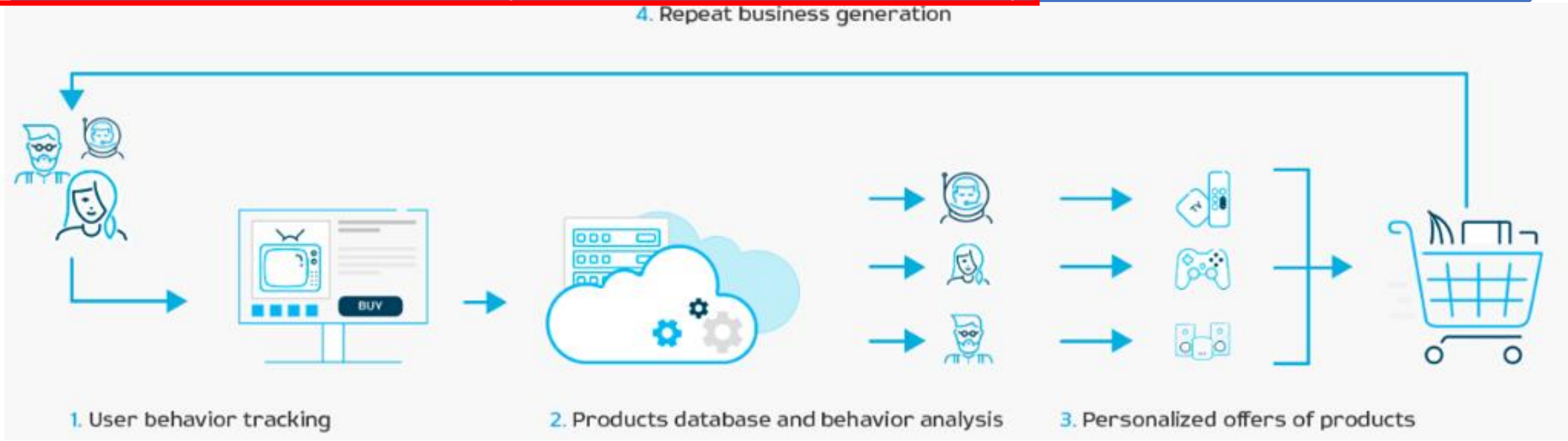
Recommendation Engine : **SmartKart**

SmartKart aims to do real-time Prediction of customized ‘Add to Cart’ recommendations based on ‘Items viewed’ by visitors and their properties on ecommerce website

Recommendation Engine : **SmartKart**

SmartKart aims to do real-time Prediction of customized ‘Add to Cart’ recommendations based on ‘Items viewed’ by visitors and their properties on ecommerce website

View | Add to Cart | Buy



Data Layout

Source	Visitor Events Data
Types	• View • AddtoCart • Buy
#Visitors	1 407 580
#Activity	2.8 million
Duration	4.5 months.

Source	Items Property
Types	• Price range • Vendor • Product type • Availability
#Property Features	800 Property for Items
Total rows	20 million

Source	Items Data
Types	Item Tree
#Item	417 053

- Source :https://www.kaggle.com/retailrocket/ecommerce-dataset/home
- Data is stored in AWS S3 - https://s3.console.aws.amazon.com/s3/buckets/smartkart/?region=us-east-2&tab=overview
- All words in text values were normalized (stemming procedure)
- All the timestamps are hashed to a sequence
- Property Data consists of snapshots for every week – means changes to any item property is captured
- All numerical values were marked with "n" char at the beginning, and have 3 digits precision

Design Layout

Amazon S3

Jupyter Notebook

Amazon SageMaker

Visitor Events Data

Items Data

Items Property Data

Data Transformation

Data Split (By Visitors)
- Train (60%)
- Test(20 %)
- Validation(20%)

Model Training

Model Deployment

Web Portal

Live Request

Live Recommendations

Predictive Scoring

Insights Dashboard

Predictive Model Details

Variable Type	Variables
Target	AddToCart(Y/N)
Predictor(800)	#times item viewed, Item ID, Price, range, Vendor,Product type, Availability, etc

Dataset	Number of rows
Training(60%)	1329253
Testing(20%)	443084
Validation(20%)	443084
Total	2215422

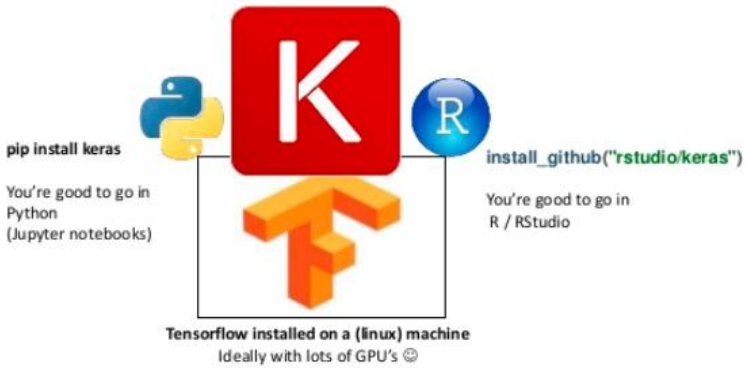
Target distribution in Training Dataset	Count
Target = 1	7758
Target = 0	1279349

Evaluation Methodology	Iteration 1 Accuracy
Confusion matrix(Classification)	82%
AUC	.84

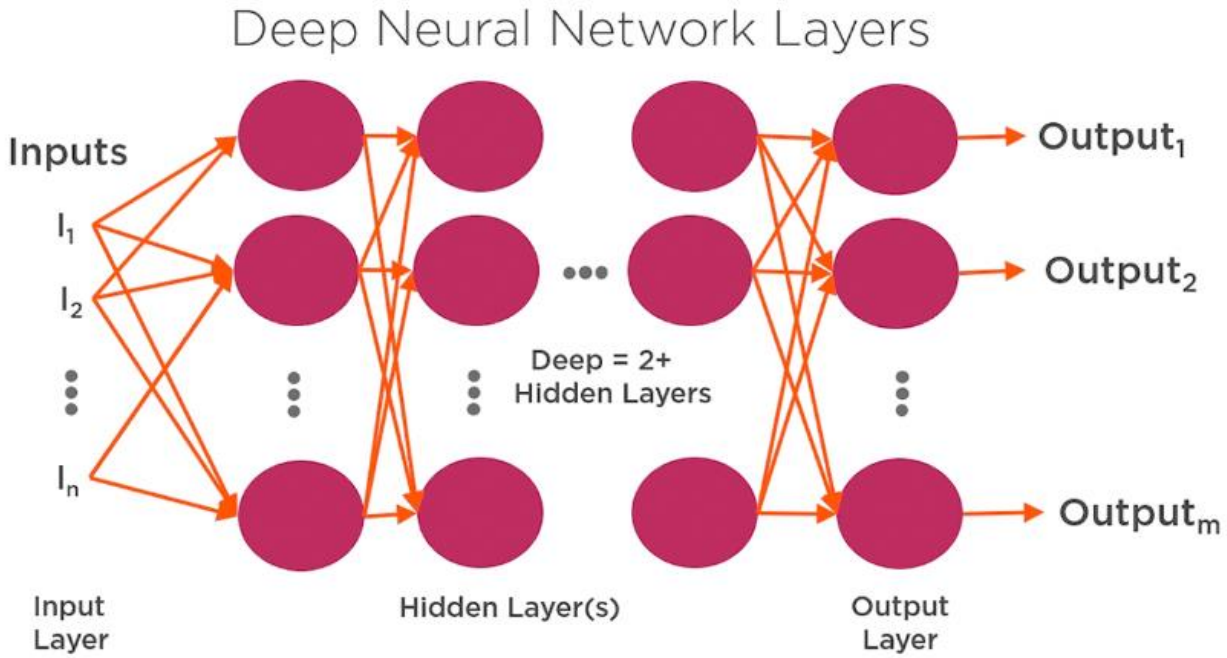
Python notebook(Code) is Available at : https://github.com/vbrahmbhatt1/ArtificialTriangle
Model results & Data available in AWS : Artificial Triangles
https://us-east-2.console.aws.amazon.com/console/home?region=us-east-2

Predictive Model Engine : Keras

- Predictive model methodology : Sequential Model Keras



https://keras.io/getting-started/



Technical Design: AWS Sage Maker

Notebook

Availability of AWS and SageMaker SDKs and sample notebooks to create training Jobs and deploy models.

Create notebook instance

Training

Train and tune models at any scale. Leverage high performance AWS algorithms or bring your own.

Training jobs

Hyperparameter tuning jobs

Inference

Create models from training jobs or import external models for hosting to run inferences on new data.

Models

Endpoints

Batch transform jobs

Learning & Next Steps

Key Learning

- Recommendation Engine
- Python /Jupyter Notebook
- Keras /Tensorflow
- AWS – Sagemaker /S3/EC2

Next steps

- Mutiple Iterations to improve accuracy
- Model Deployment
- Connecting Web Portal to Model in order to enable live scoring

References

- https://arxiv.org/abs/1511.06939