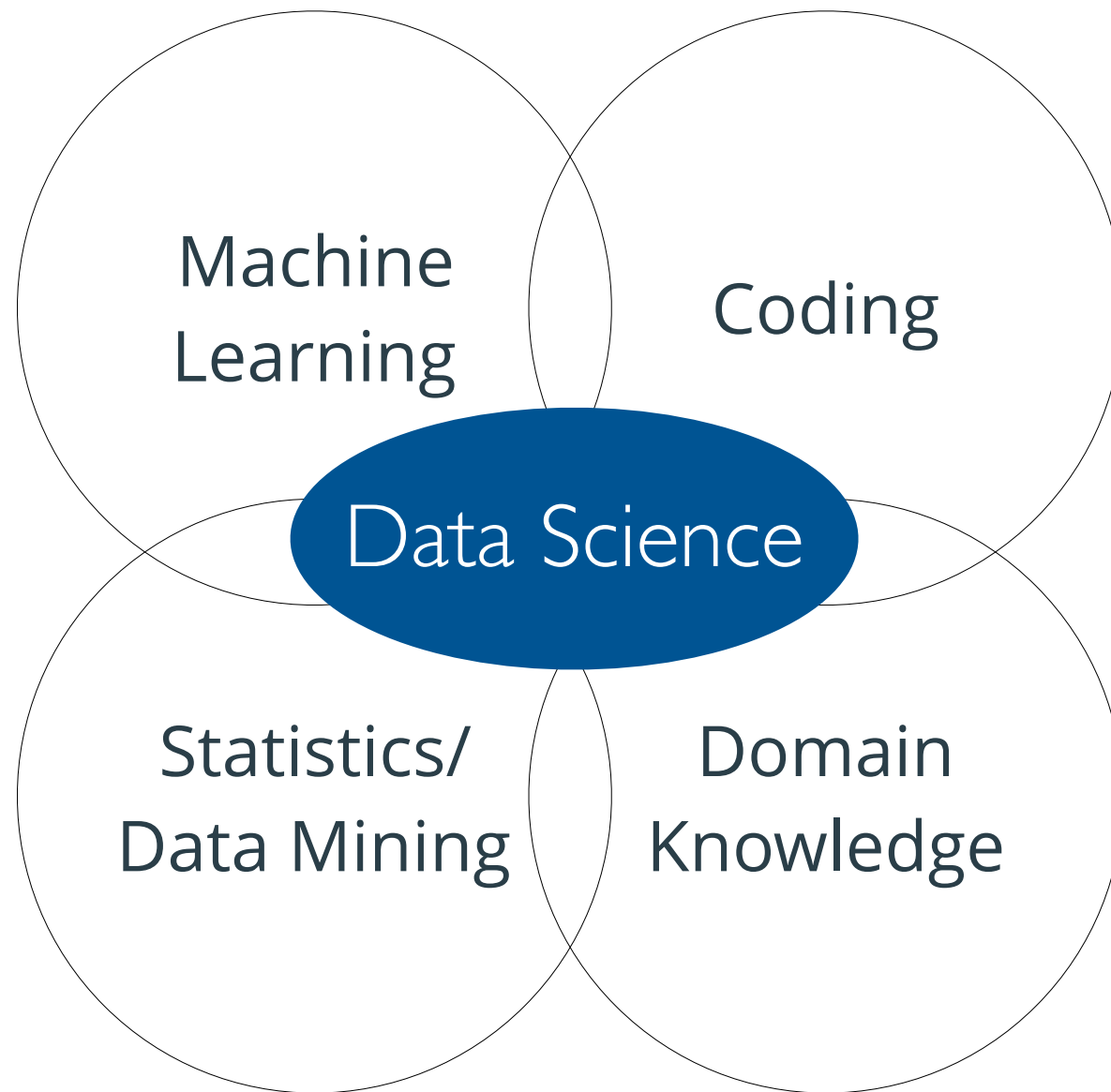


BOOSTING PRODUCT CATEGORIZATION WITH MACHINE LEARNING

Amadeus Magrabi

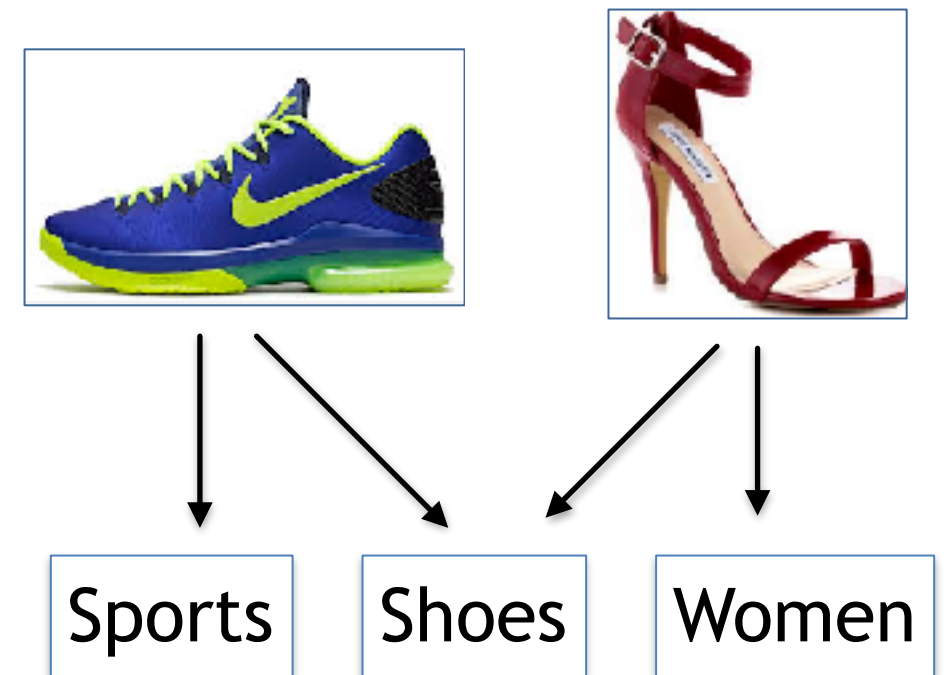
@amadeusmagrabi

Background



Motivation

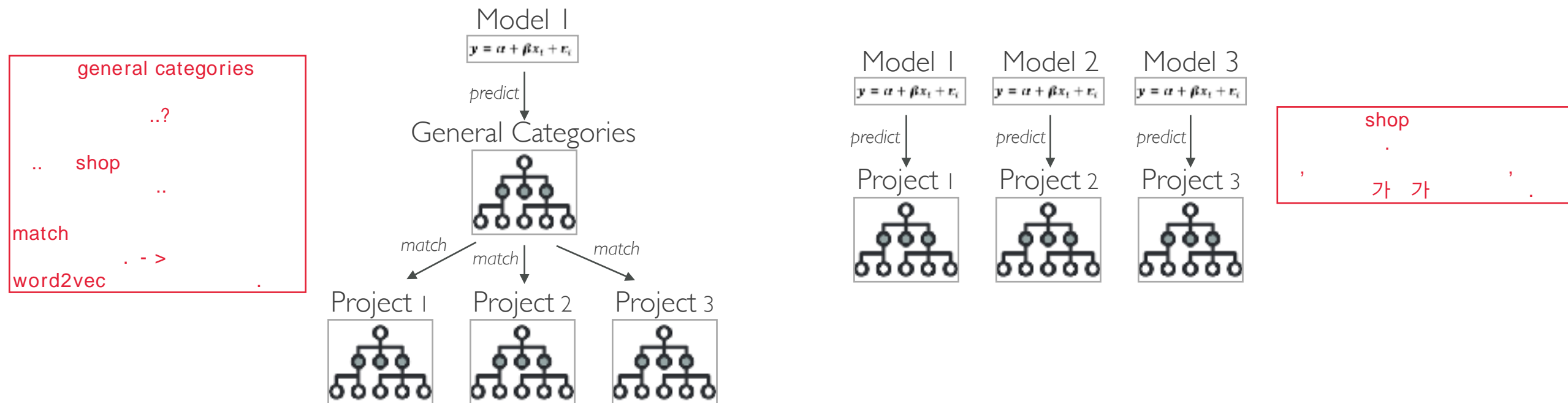
- **Product categories** define the structure of commerce platforms
- To facilitate **user navigation**, categories need to be **accurate, logical, and consistent**
- **Problem:** A lot of manual work to maintain a clean and accurate category structure
- **Goal:** Use machine learning to automatically recommend categories for products



Approach

, , 가 , 가 .

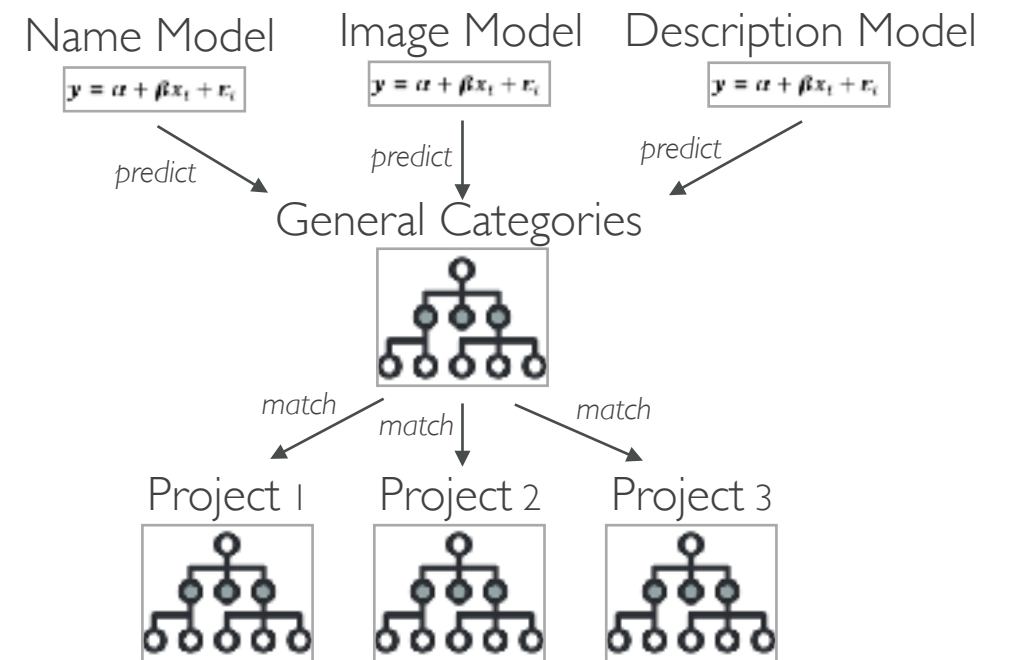
- **Modeling approach:** one general model vs. multiple project-specific models



- Advantages of **general model**:
 - Can also be applied to projects with little data
 - Better control of model building (e.g. handle imbalanced data)
 - Can also recommend new categories
- Advantages of **multiple project-specific models**:
 - Better accuracy for projects with very specific categories
 - No category matching procedure necessary

Approach

- Focus on product **names**, **images** and **descriptions**
 - carry most information
 - available for most products
- Other product attributes like **price**, **size**, **color**, etc. not included so far
- Language: **python**

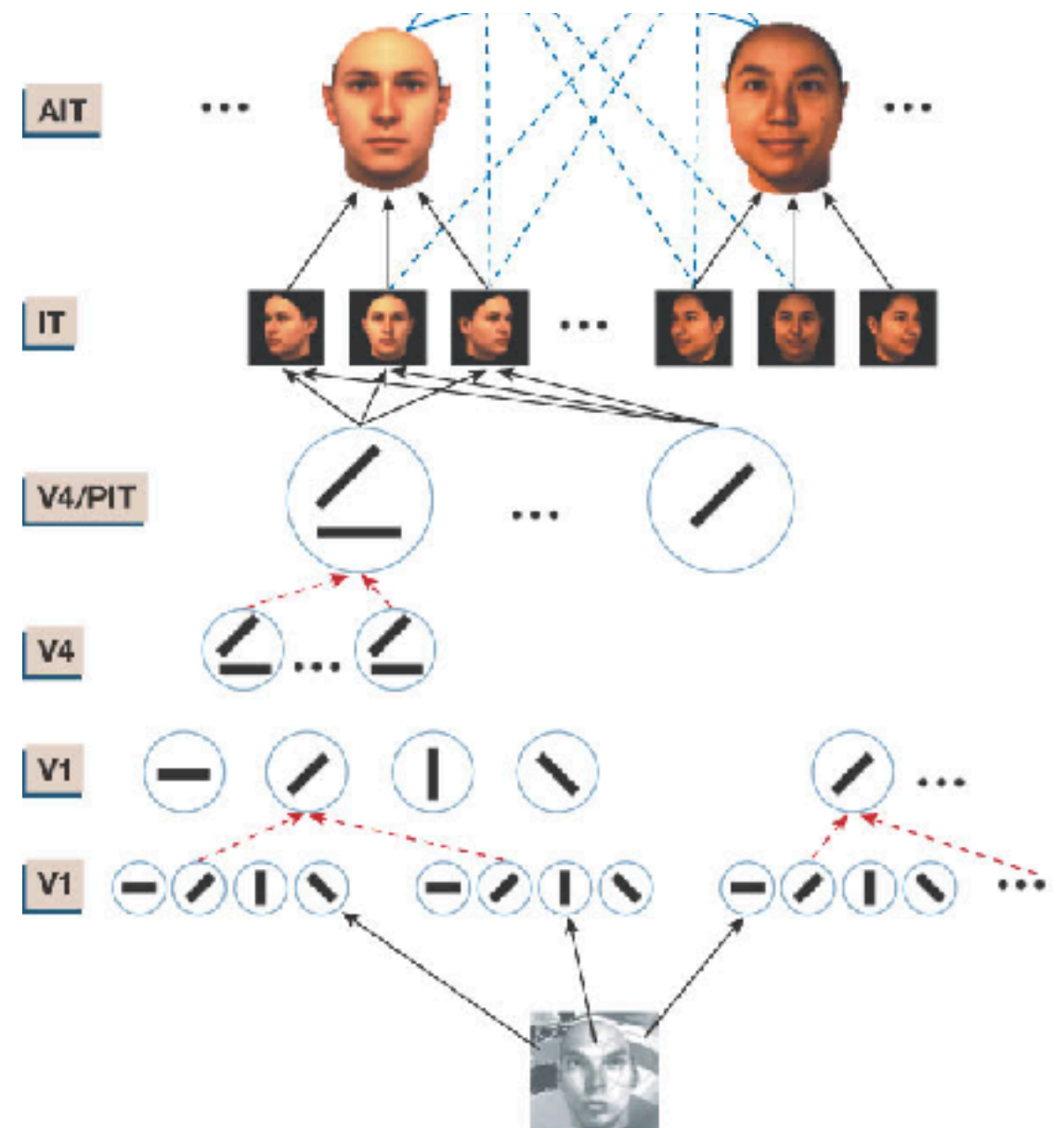
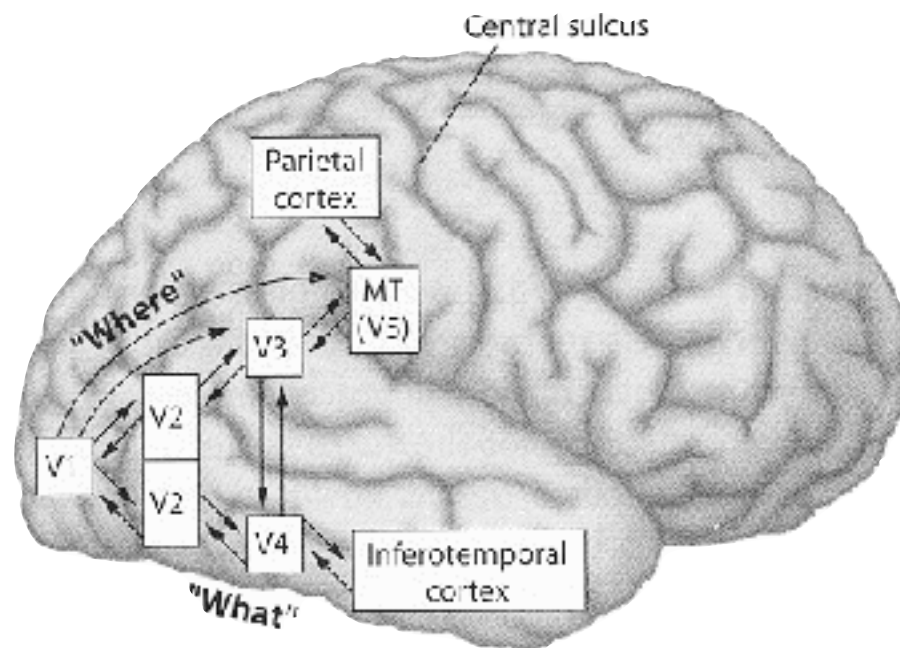


Product Categories

5-Pocket Alcohol Animal Food Animals Ankle Boots Antennas Apples Armchairs Art Baby Backpacks Bags
Baguettes Baking Ballerinas Balls Bananas Bandanas Barbecue Batteries Beanies Beans Beef Beer Belts
Beverages Bikes Bikinis Birthday Blankets Board Games Books Boots Bottles Bow Ties Bowls Boxershorts
Boxes Boys Bras Bread Brooms Buckets Burgers Business Butter Cables Cake Candles Candy Cans Card
Games Cargo Pants Carpets Carrots Cars Casual Cds Cereals Chairs Champagne Chargers Cheese Chicken
Children Chili Chinos Chocolate Christmas Cigarettes Cleaning Clothing Cloths Clutches Coats Cocoa
Coffee Coffee Machines Computer Consoles Convenience Foods Cookies Cooking Oil Cucumbers Cups
Curtains Cutlery Deodorants Detergents Diapers Dish Washer Dishes Drawing Dresses Drinks Dvds Eggs
Fashion Fast Food Figurines Fish Flour Flowers Food Footwear Formal French Fries Fruits Furniture Games
Girls Glass Glasses Gloves Gold Gum Hair Hammers Handbags Hats Headbands Headphones Headwear
Helmets High Heels Honey Hoodies Hygiene Ice Ice Cream Jackets Jars Jeans Jewelry Juices Keyboards
Kitchen Knives Lace-Up Shoes Lamps Laptops Leather Lettuce Lights Lipsticks Loafers Long Sleeves Lotions
Magazines Makeup Mascara Meat Medicine Melons Men Metal Military Milk Mittens Monitors Motorcycles
Mushrooms Music Nail Polish Nails Necklaces Noodles Nuts Office Supplies Oranges Overalls Paintings Pans
Pants Paper Parkas Party Pasta Pencils Pens Pepper Perfumes Phones Pillows Pizza Plants Plates Polo Shirts
Porcelain Pork Posters Potatoes Pots Power Supply Prawn Printers Projectors Pullovers Pumps Radios
Rainwear Razors Reading Receivers Refrigerators Rice Rings Routers Salads Salmon Salt Sandals Sandwiches
Sauces Sausages Scarves Scissors Screwdrivers Screws Sculptures Shampoo Shaving Shoes Short Sleeves
Shorts Shoulder Bags Shrimps Silver Skin Cream Skirts Slips Smartphones Snacks Sneakers Socks Soft Drinks
Soups Soy Speakers Spices Spoons Sports Steaks Stickers Suits Summer Sunglasses Sunscreen Sweat Jackets
Sweat Pants Sweatshirts Sweets Swimsuits Swimwear T-Shirts Tables Tablets Tank Tops Tea Television Ties
Tissues Toasters Toiletries Tomatoes Tools Toothbrushes Toothpaste Tops Towels Toys Trousers Tuna Undershirts
Underwear Vases Vegetables Vegetarian Vests Video Games Vitamins Wallets Wardrobe Water Wedding Wine
Winter Women Wood Wraps Yoga Yoghurt

Model for Product Images

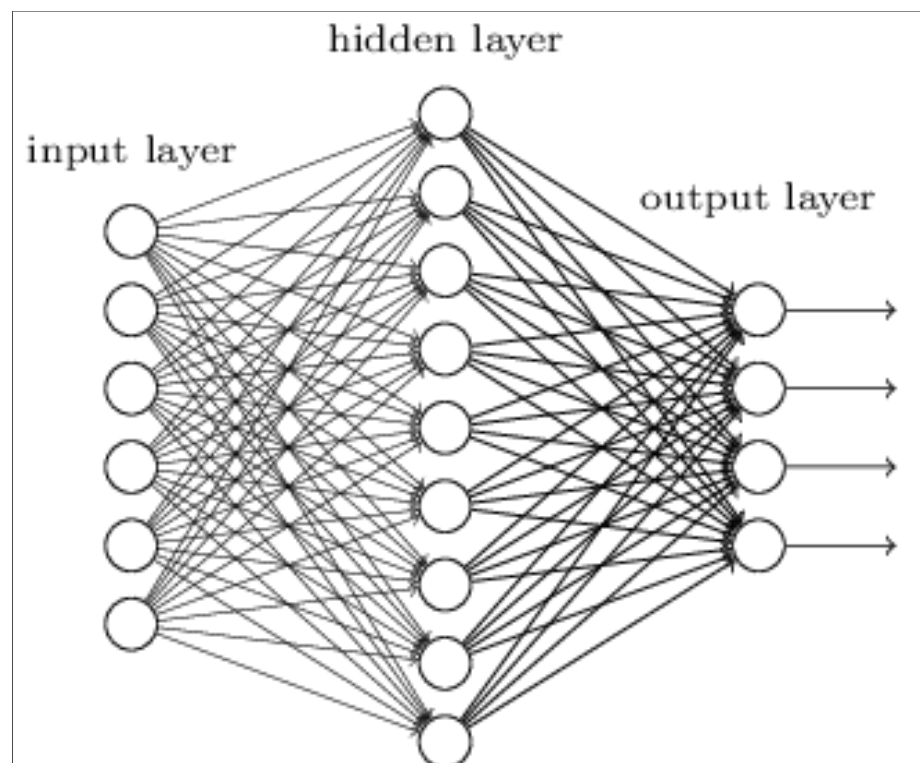
- **Model:** Neural Network (“Deep Learning”)
 - Most successful machine learning model for image classification tasks
 - Explosive use in recent years, mainly because of more available computing power
- **Based on mechanisms in the brain:**
 - Idea of building complex representations by connecting simple representations



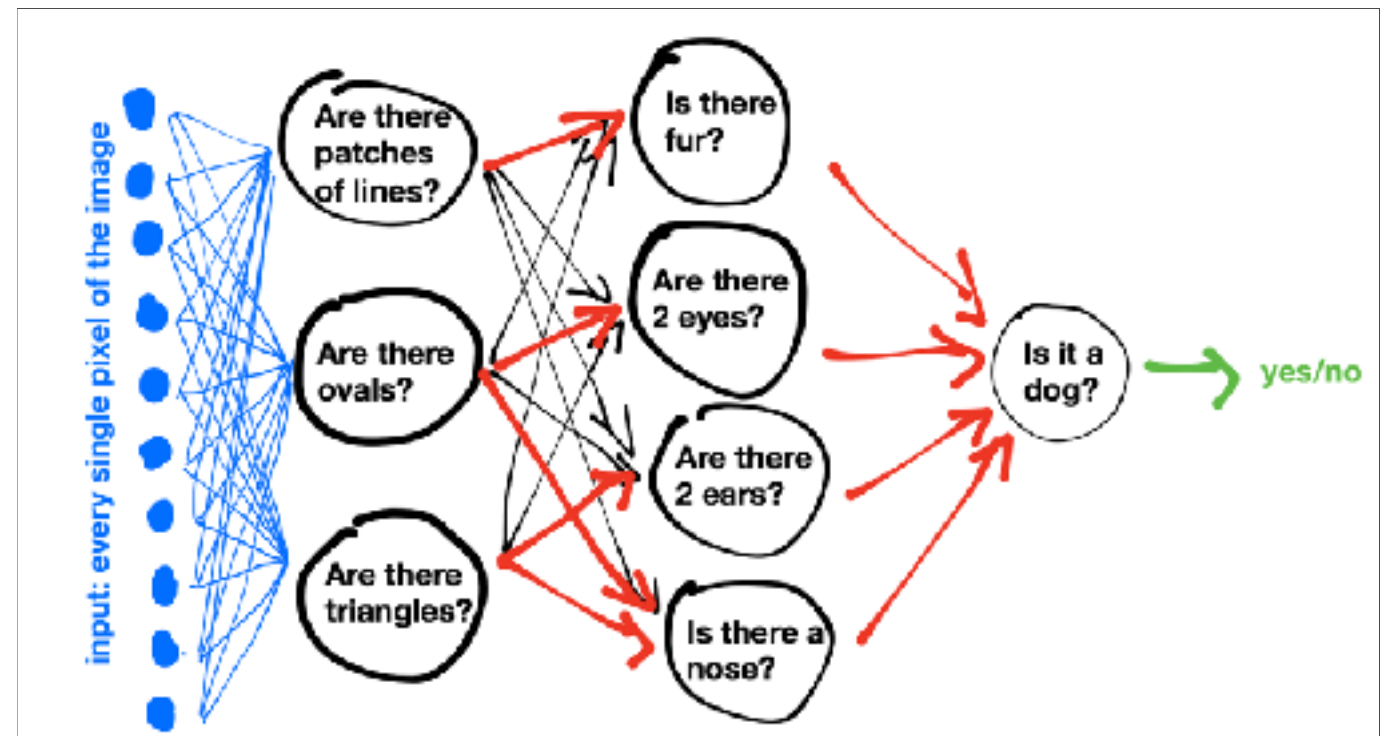
Model for Product Images

Basics of Neural Networks:

Neural Network

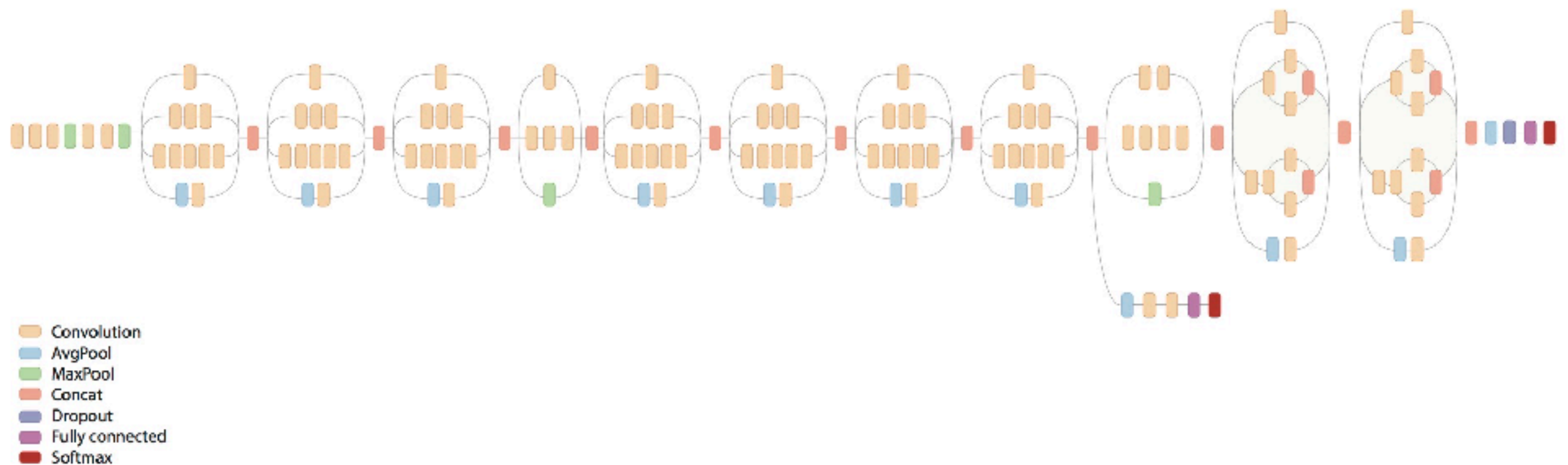


Convolutional Neural Network



Model for Product Images

- **Model training:**
 - Transfer learning on image recognition network [Inception v3](#) [library: *TensorFlow*]
 - Computed on [Google Cloud](#)



Model for Product Names

Examples:

"Mens Heavyweight 6.1-ounce, 100% cotton T-Shirts in Regular, Big and Tall Sizes"

"Gala Apples Fresh Fruit, 3 LB Bag"

"Carhartt Men's Maddock Pocket T-Shirt Size M"

"Samsung SM-G900V - Galaxy S5 - 16GB Android Smartphone Verizon + GSM - Black"

- **Preprocessing Steps:** [library: *spacy*, *re*]

consistent form
가 .

- *spellchecker* (smartwathc → smartwatch)
- *translation* (German → English)
- *tokenization* (sentences → words)
- *normalization* (lowercasing, remove special characters)
- *lemmatization* (apples → apple)
- *phrasing* (louis vuitton → louis_vuitton)
- *word removal* (stop words, blacklist)

phrasing
, . phrasing

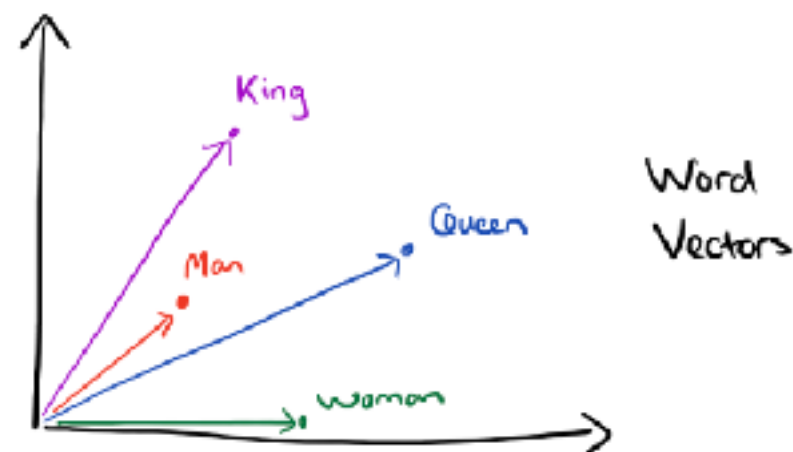
Model for Product Names

- **Vectorization methods:** (*text* \rightarrow *numbers*)
 - bag-of-words:
 - Simple approach, but sparse representation and blind to context


the dog is on the table

0	0	1	1	0	1	1	1
are	cat	dog	is	now	on	table	the

- **word2vec:** Trains two-layer neural network to form a dense and context-sensitive representation



Model for Product Names

- **word2vec model training:**
 - trained on dataset from Google news [*library: gensim*]
 - lexicon of 3 million words, word vectors with n=300
- **Prediction Models:** [*library: scikit-learn*]
 - Logistic Regression 
 - Naive Bayes
 - Random Forest
 - XGBoost
 - Support Vector Machine

Model for Product Descriptions

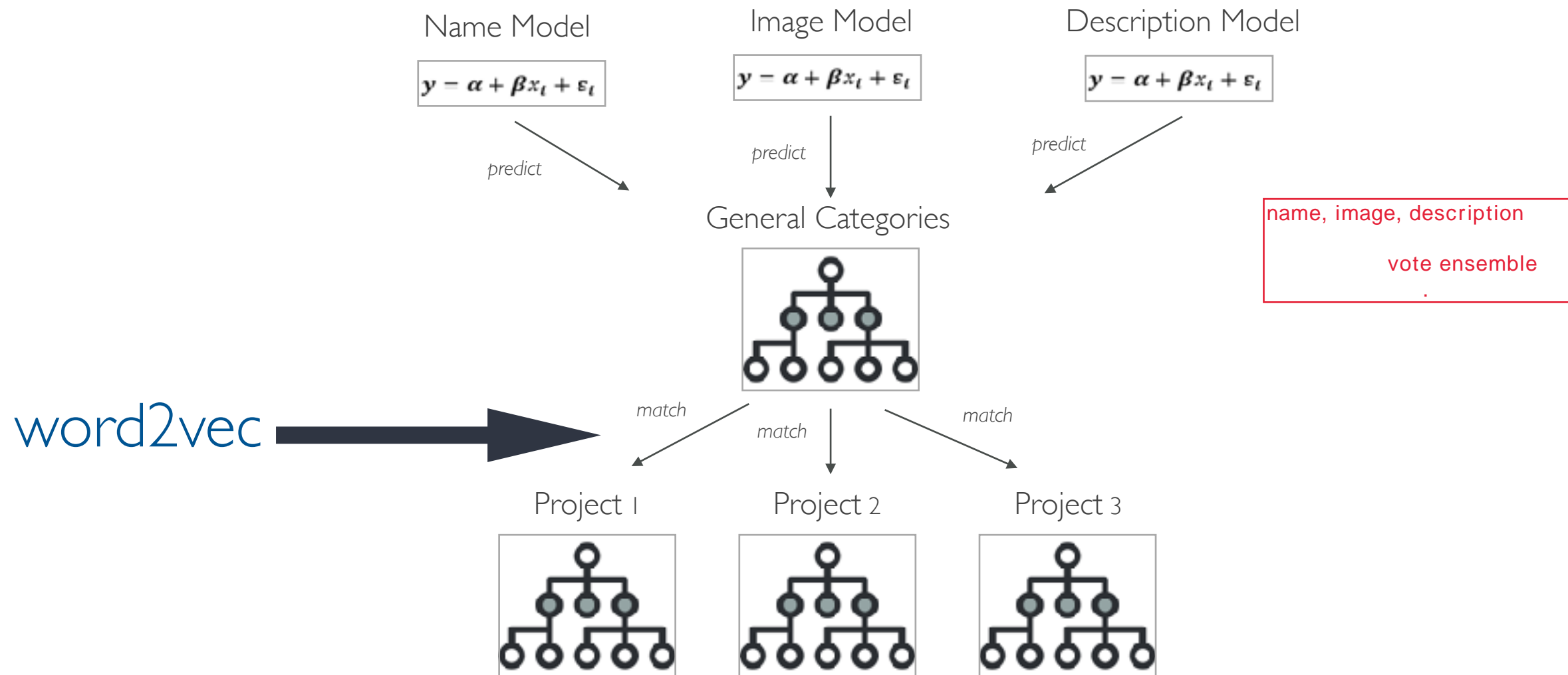
Model for Product Names



same approach

Model for Product Descriptions

Project-Matching

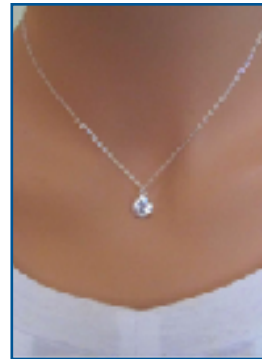


REST API



```
(ml-services) MacBook-Pro-2:~ amagrabi$ curl -G -s \  
> -d img_url=https://s3-eu-west-1.amazonaws.com/commercetools-maximilian/products/079474_1_large.jpg \  
> -d num=3 \  
> -d category_type=general \  
> http://127.0.0.1:8000/categorizer | jq .  
{  
  "results": [  
    {  
      "category": "Formal",  
      "confidence": 0.56782  
    },  
    {  
      "category": "Suits",  
      "confidence": 0.02535  
    },  
    {  
      "category": "Business",  
      "confidence": 0.02203  
    }  
  ],  
  "input": {  
    "image": "https://s3-eu-west-1.amazonaws.com/commercetools-maximilian/products/079474_1_large.jpg"  
  },  
  "count": 3,  
  "code": 200  
}
```

REST API



```
(ml-services) MacBook-Pro-2:~ amagrabi$ curl -G -s \  
> -d img_url=https://s-media-cache-ak0.pinimg.com/736x/5c/23/cd/5c23cdb8b7cecf9f0932596f3442c00.jpg \  
> -d name=Mother%27s%20Day%20Gift%20-%20Infinity%20Pendant%20Name%20Necklace%20in%20Sterling%20Silver%20-%20Custom%20Made%20with%20Any%20Name%21 \  
> -d num=3 \  
> -d category_type=general \  
> http://127.0.0.1:8000/categorizer | jq .  
{  
  "results": [  
    {  
      "category": "Necklaces",  
      "confidence": 0.17829  
    },  
    {  
      "category": "Jewelry",  
      "confidence": 0.058  
    },  
    {  
      "category": "Silver",  
      "confidence": 0.0336  
    }  
  ],  
  "input": {  
    "name": "Mother's Day Gift - Infinity Pendant Name Necklace in Sterling Silver - Custom Made with Any Name!",  
    "image": "https://s-media-cache-ak0.pinimg.com/736x/5c/23/cd/5c23cdb8b7cecf9f0932596f3442c00.jpg"  
  },  
  "count": 3,  
  "code": 200  
}
```


REST API

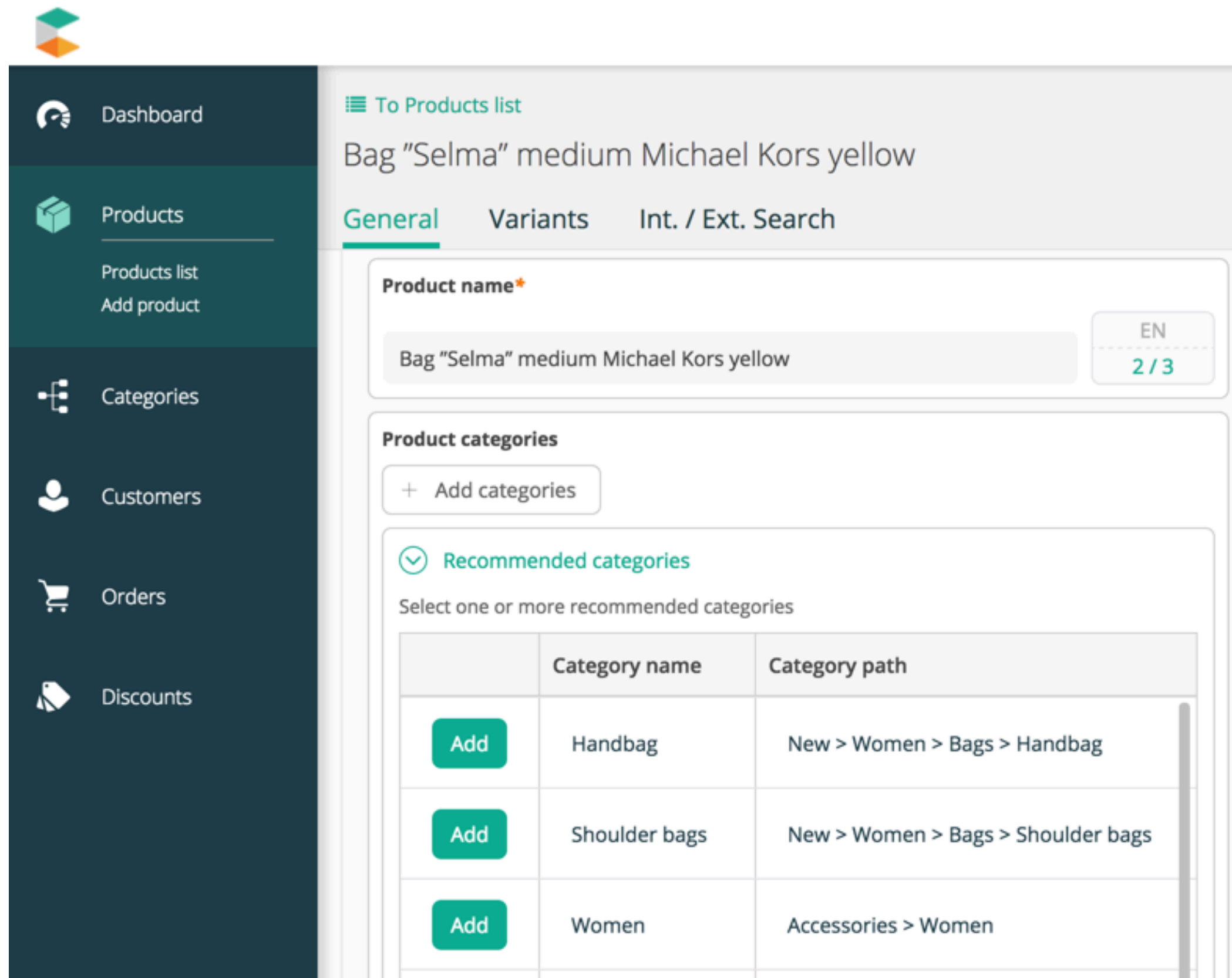
```
(ml-services) MacBook-Pro-2:~ amagrabi$ curl -G -s \  
> -d name=McDonalds \  
> -d num=3 \  
> -d category_type=general \  
> http://127.0.0.1:8000/categorizer | jq .  
{  
  "results": [  
    {  
      "category": "French Fries",  
      "confidence": 0.1084  
    },  
    {  
      "category": "Burgers",  
      "confidence": 0.0689  
    },  
    {  
      "category": "Fast Food",  
      "confidence": 0.03858  
    }  
  ],  
  "input": {  
    "name": "McDonalds"  
  },  
  "count": 3,  
  "code": 200  
}
```

REST API



```
(ml-services) MacBook-Pro-2:~ amagrabi$ curl -G -s \  
> -d prod_id=22789cbb-28bc-46c4-876a-d5d3f19b330f \  
> -d access_token=${ACCESS_TOKEN} \  
> -d project_key=${PROJECT_KEY} \  
> -d host=EU \  
> -d staged=false \  
> -d num=1 \  
> -d category_type=project \  
> http://127.0.0.1:8000/categorizer | jq .  
{  
  "results": [  
    {  
      "confidence": 0.3898,  
      "category": {  
        "path": "Accessories > Men > Sunglasses",  
        "id": "4ad97691-53b5-401b-bd22-0bb91c0e44c7",  
        "name": {  
          "it": "Occhiali da sole",  
          "en": "Sunglasses",  
          "de": "Sonnenbrillen"  
        },  
      },  
    },  
  ],  
  "input": {  
    "productId": "22789cbb-28bc-46c4-876a-d5d3f19b330f",  
    "name": "Sunglasses Spektre grey",  
    "image": "https://s3-eu-west-1.amazonaws.com/commercetools-maximilian/products/081969_1_large.jpg"  
  },  
  "count": 1,  
  "code": 200  
}
```

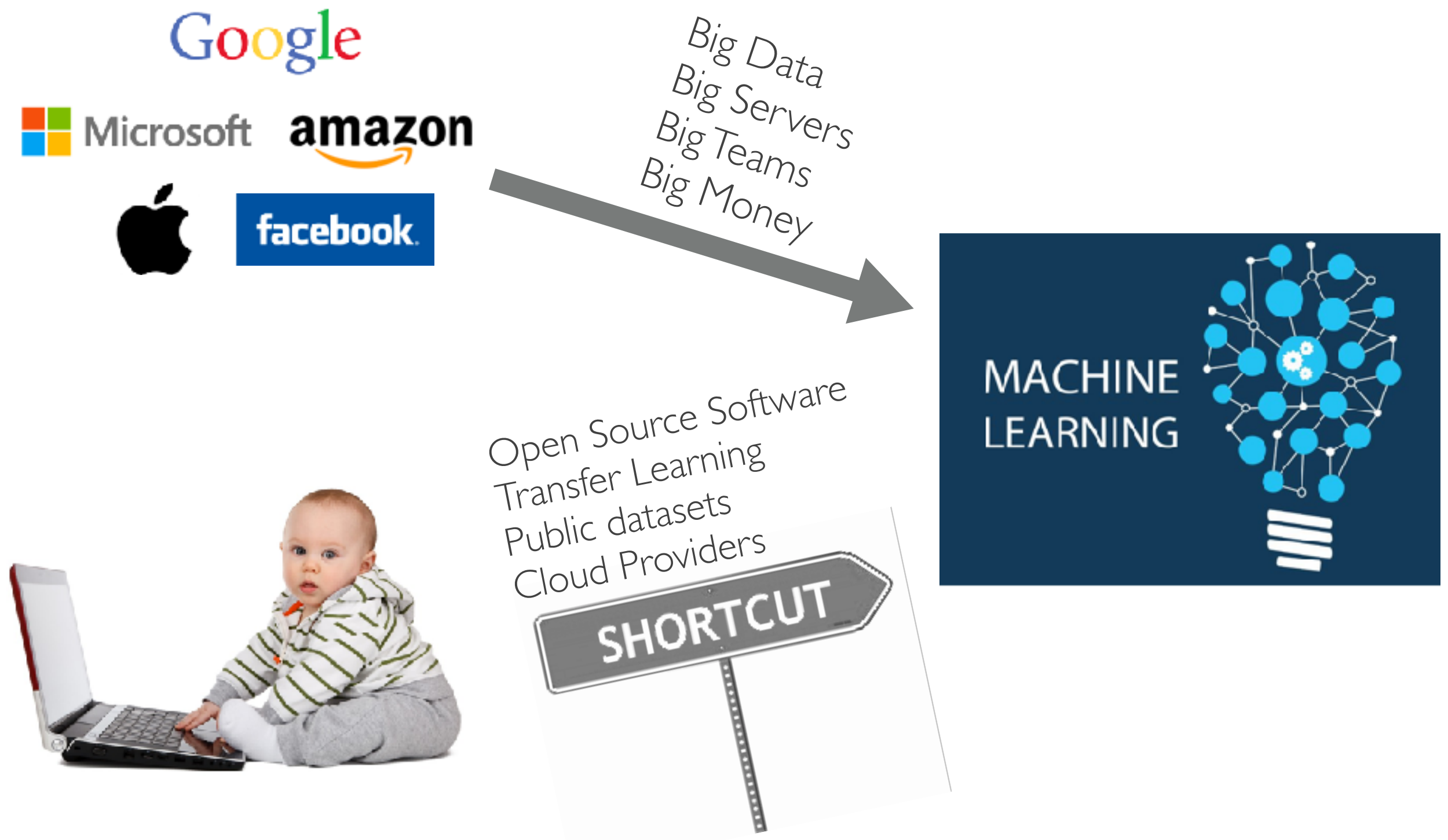
UI in Merchant Center



The screenshot displays the Merchant Center interface. On the left is a dark sidebar with navigation links: Dashboard, Products (highlighted), Categories, Customers, Orders, and Discounts. The main content area shows the product editing page for 'Bag Selma'. At the top, there's a link 'To Products list' and the product title 'Bag Selma medium Michael Kors yellow'. Below the title are tabs for 'General' (active), 'Variants', and 'Int. / Ext. Search'. The 'General' tab contains a 'Product name' field with the same text and a language selector set to 'EN' (2/3). Below this is a 'Product categories' section with an 'Add categories' button. A 'Recommended categories' section follows, with a checkmark icon and the text 'Select one or more recommended categories'. It contains a table with three rows of recommended categories, each with an 'Add' button.

	Category name	Category path
<button>Add</button>	Handbag	New > Women > Bags > Handbag
<button>Add</button>	Shoulder bags	New > Women > Bags > Shoulder bags
<button>Add</button>	Women	Accessories > Women

Final Remark: Data Science in Big vs. Small Companies





Thank you!

Amadeus Magrabi

 @amadeusmagrabi

amadeus.magrabi@commercetools.com

www.commercetools.com