

# Classification

*Work Like Tomorrow.™*

**KOFAX**

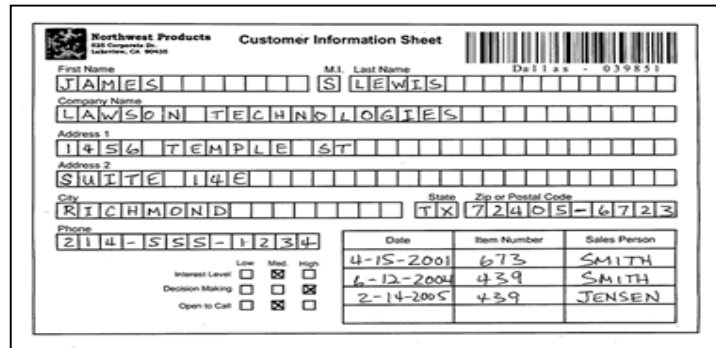
# What is Classification?



- The ability to confidently recognize a document type by its layout, content, or both

# Recap - Structured, Semi-structured, Unstructured

- Kofax TotalAgility 7 includes modules and activities that work with Kofax Transformation Designer to provide advanced document classification, separation, and extraction for structured, semi-structured and unstructured documents



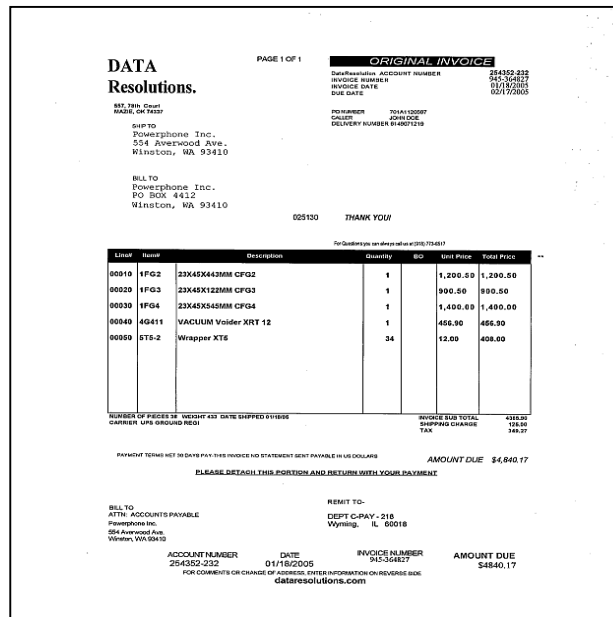
**Northwest Products** Customer Information Sheet

First Name: JAMES M.I. Last Name: S LEWIS  
 Company Name: LAWSON TECHNOLOGIES  
 Address 1: 1456 TEMPLE ST  
 Address 2: SUITE 14E  
 City: RICHMOND State: TX Zip or Postal Code: 72405-6723  
 Phone: 214-555-1234

Interest Level: ☐ Low ☒ Med ☐ High  
 Decision Making: ☐ ☒ ☐  
 Open to Call: ☐ ☒ ☐

Date	Item Number	Sales Person
4-15-2001	673	SMITH
6-12-2004	439	SMITH
2-14-2005	439	JENSEN

Structured



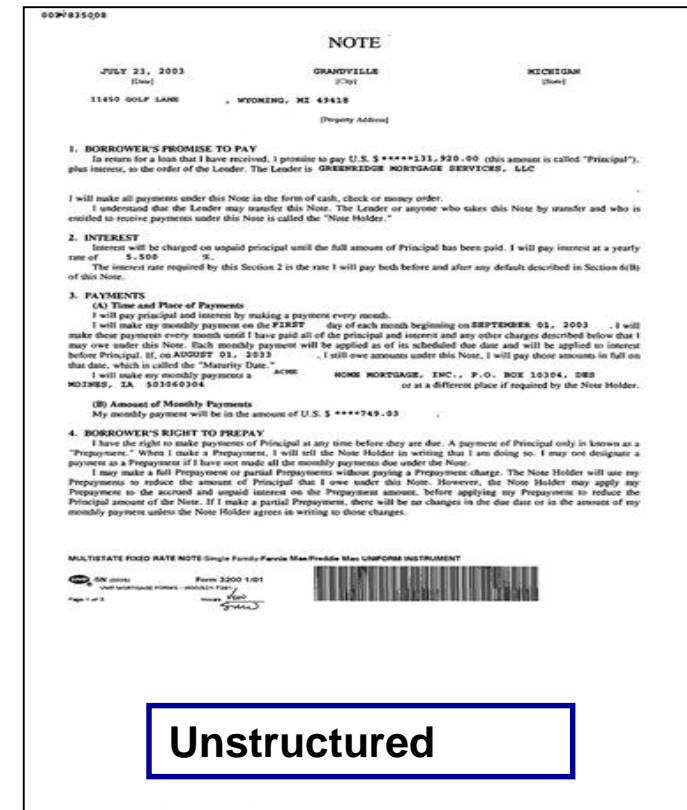
**DATA Resolutions.** ORIGINAL INVOICE

DATE: 01/18/2005 INVOICE NUMBER: 915-34827 AMOUNT DUE: \$4840.17

Line	Item	Description	Quantity	EO	Unit Price	Total Price
00010	1FG2	23X45X443MM CFQ2	1		1,200.50	1,200.50
00020	1FG3	23X45X122MM CFQ3	1		900.50	900.50
00030	1FG4	23X45X343MM CFQ4	1		1,400.00	1,400.00
00040	4G41	VACUUM Vender XRT 12	1		456.90	456.90
00050	ST-2	Wrapper XTS	24		12.00	408.00

AMOUNT DUE: \$4,840.17

Semi-structured



**NOTE**

JULY 23, 2003 (Date) GRANDVILLE (City) MICHIGAN (State)

11450 GOLF LAKE, WYOMING, MI 49418 (Property Address)

**1. BORROWER'S PROMISE TO PAY**  
 In return for a loan that I have received, I promise to pay U.S. \$ \*\*\*\*131,920.00 (this amount is called "Principal"), plus interest, to the order of the Lender. The Lender is: GREENRIDGE MORTGAGE SERVICES, LLC

**2. INTEREST**  
 Interest will be charged on unpaid principal until the full amount of Principal has been paid. I will pay interest at a yearly rate of 5.500 %.

**3. PAYMENTS**  
 (A) Time and Place of Payments:  
 I will pay principal and interest by making a payment every month. I will make my monthly payment on the FIRST day of each month beginning on SEPTEMBER 01, 2003. I will make these payments every month until I have paid all of the principal and interest and any other charges described below that I may owe under this Note. Each monthly payment will be applied as of its scheduled due date and will be applied to interest before Principal. If, on AUGUST 01, 2013, I still owe amounts under this Note, I will pay those amounts in full on that date, which is called the "Maturity Date."

**4. BORROWER'S RIGHT TO PREPAY**  
 I have the right to make payments of Principal at any time before they are due. A payment of Principal only is known as a "Prepayment." When I make a Prepayment, I will tell the Note Holder in writing that I am doing so. I may not designate a payment as a Prepayment if I have not made all the monthly payments due under the Note.

Unstructured

# Analyze Documents



- Classification
  - Get a large stack of paper and try to sort it
  - Have you recognized it by a specific layout?
    - ➔ Layout classification
  - Have you read the whole document?
    - ➔ Content classification (Adaptive Feature classifier)
  - Have you read just a phrase or a word?
    - ➔ Content classification (Instruction)

# Classification



- Transformation Designer supports following classification methods
  - Layout classification
  - Content classification “Adaptive feature classifier”
  - Content classification “Instructions”

# Classification

- Select **Layout**, **Content**, or **Both** methods for Classification
- Provide **Training Samples** for each Document Type
- **Train** the Project
- **Test** Classification at Design Time using **Test Document Sets**
- Examine **Classification Confidence**
- Test Classification at **Run Time** by Creating a Job



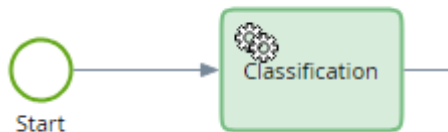
In the Training Example we will use Layout Classification.  
(Content Classification is covered in other courses)

# Configuring Classification

- Open the **Classification Group** in the Transformation Designer
- Configure **Classification** at the Project Level
  - Layout only (in this example)
- Obtain good quality **samples** (e.g. at least 1, 300 dpi, black and white)
  - Open the Samples Document Set
  - Select the Class (Document Type)
  - Add the sample(s) to the Classification Set
  - Repeat for each Class
- **Train** the Project
- **Test** on a Document Test Set

# Classification Activity

- Automated activity that executes Classification project information
- Requires an Input Folder
- Requires Classification Group (which can be inherited from the Process)
- Separation Profile only required if separation will take place during classification rather than on scan



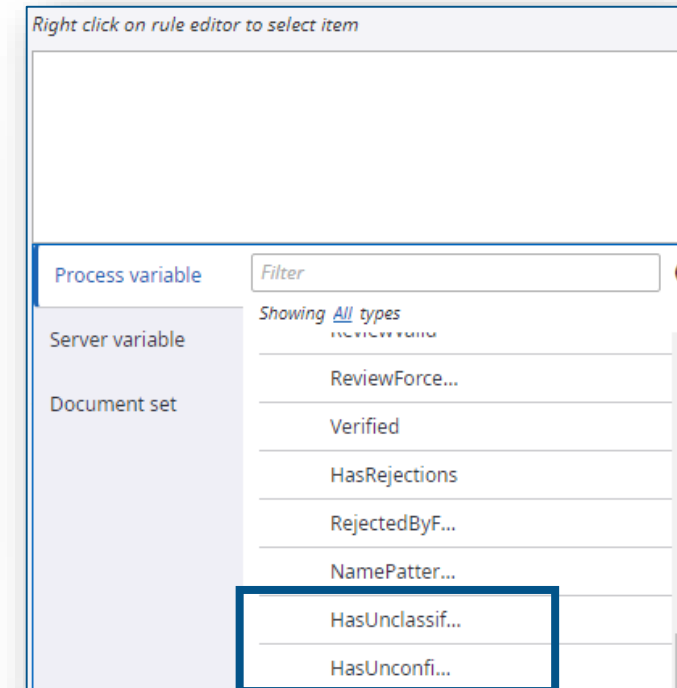
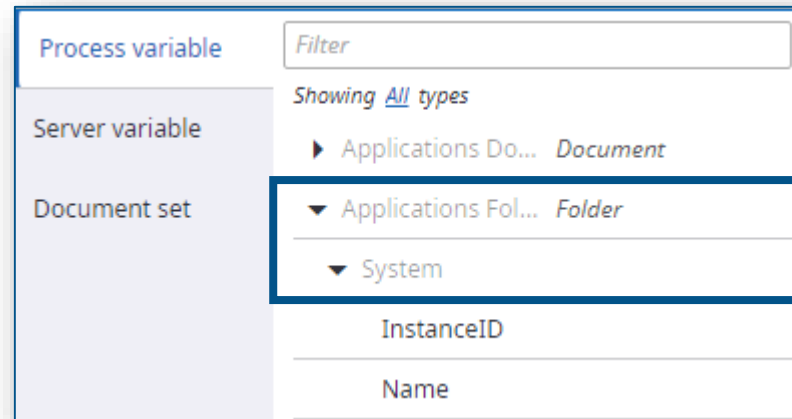
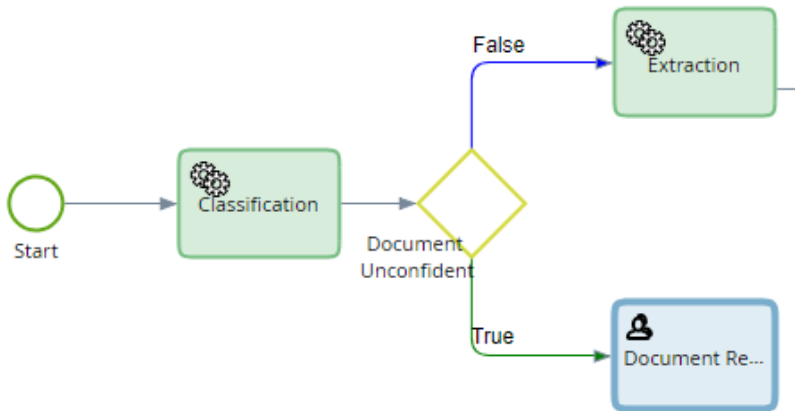
The screenshot shows the configuration window for the 'Activity - classification'. The window has a sidebar with icons for settings, users, configuration, a timer, a folder, a classification, and a flag. The main area contains the following fields:

- Name:** A text box containing 'Classification'.
- Type:** A dropdown menu with 'Classification' selected.
- Configuration:**
  - Input folder:** A dropdown menu with 'Applications Folder (Process variable) ×' selected.
  - Separation profile:** An empty dropdown menu.
- Classification:**
  - Inherit from process:** A checkbox that is checked.



# What is Document Review?

- Optional manual activity to **review classification** results
- Document Review can be bypassed by checking the folder system fields: HasUnconfidentlyClassifiedDocument returns a Boolean (True/False) value
- HasUnconfidentlyClassifiedDocument returns a Boolean (True/False) value
- You can use a decision activity to test its value after classification and bypass Document Review if required



# Test Classification at Run Time

- You can test classification at run time by creating a job in the TotalAgility Workspace
- The **Document Review** manual activity allows users to review (and modify) classification results

The screenshot displays the TotalAgility Workspace interface. On the left, the 'Work Queue' sidebar shows a tree view with 'My Queries', 'Shared Queries', 'Alerts', and 'All Activities (default)'. The main area is divided into three panes. The top pane shows a table with columns 'Activity Name' and 'Process', containing the entry 'Document Review' and 'Capture Applications'. The bottom-left pane, titled 'Navigator', shows a tree view with 'Applications (2)' containing two items: '1: Standard (1)' with a confidence of 95.0% and '2: Standard (1)' with a confidence of 91.8%. The bottom-right pane, titled 'Page 1 of 1', shows a document preview of a 'KOFAX' form titled 'XYZ Bank - Personal Loan Application'. The form includes a section '1. Personal details – main applicant' with a 'Title' field where 'Mr' is selected.

Activity Name	Process
Document Review	Capture Applications

Name	Type	Confidence
Applications (2)		
1: Standard (1)	Standard	95.0%
2: Standard (1)	Standard	91.8%

Page 1 of 1

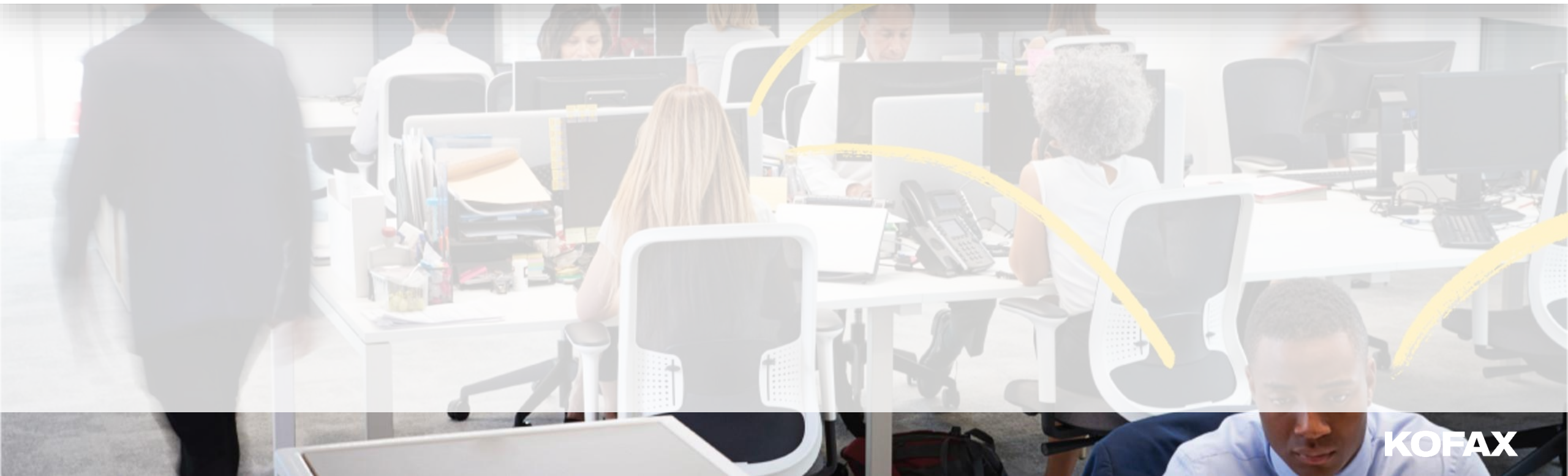
KOFAX

XYZ Bank - Personal Loan Application

1. Personal details – main applicant

Title Mr ☒ Mrs ☐ Miss ☐ Ms ☐ Other ☐

# Classification Online Learning

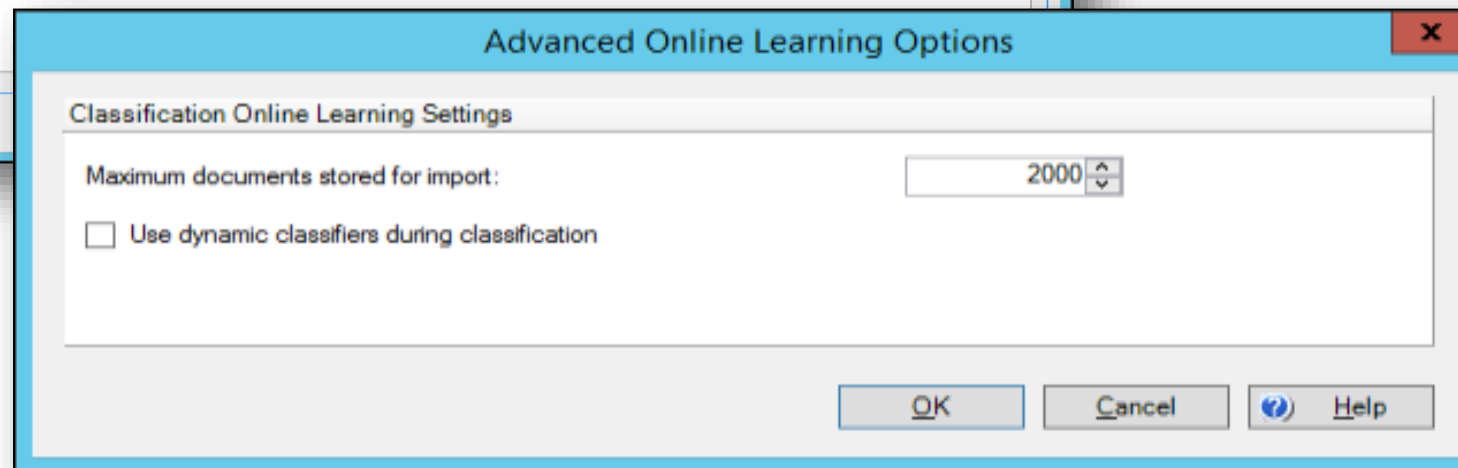
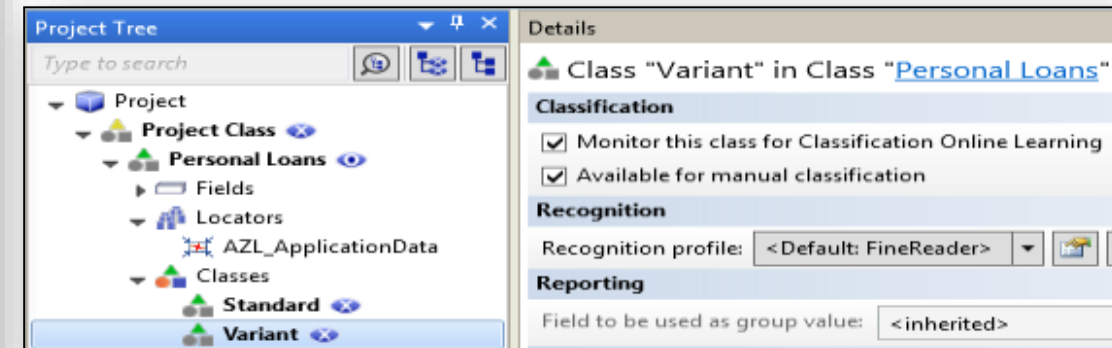
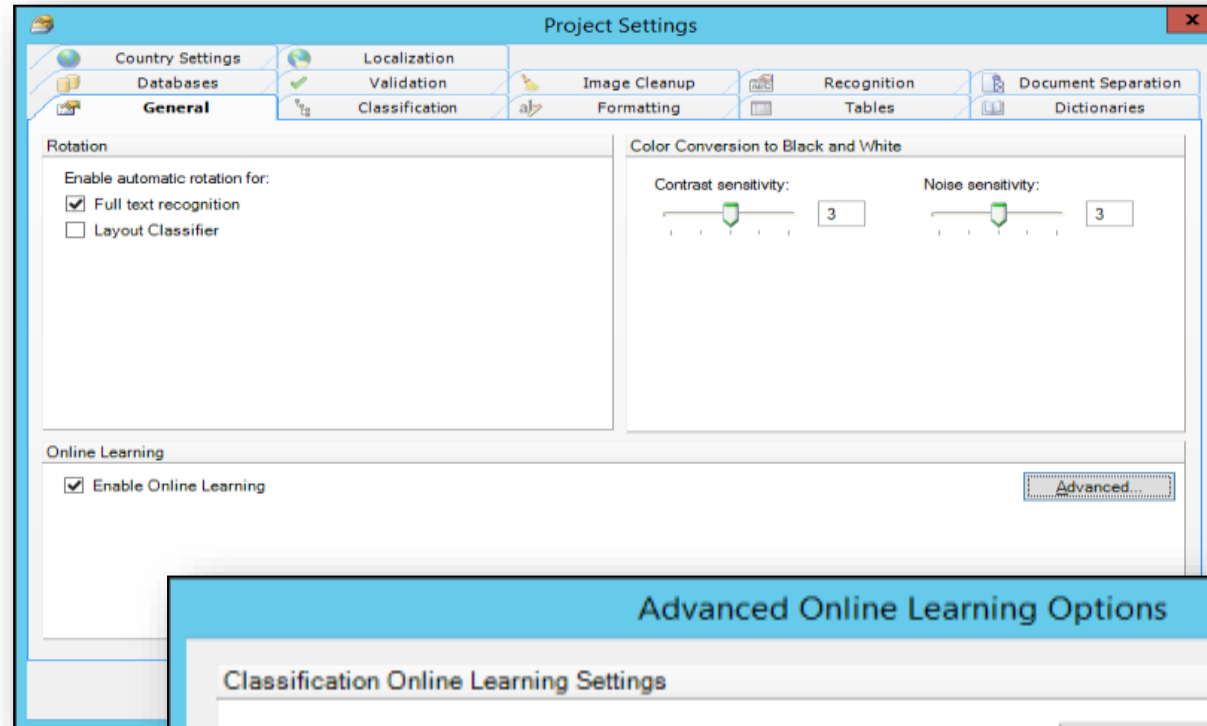


# Classification Online Learning

- When a new layout is encountered, the document reviewer can manually change its type. The next time this document is encountered it will automatically be classified
  - Online Learning must be enabled in the Classification project
  - Monitor this class (document type) for Classification Online Learning must be enabled in the Extraction project
  - An operator can also type the first few characters of a document type to filter the document type list (Auto Complete)

Name	Type	Confidence
Applications (2)		
1: Standard (1)	Variant x	95.0%
2: Standard (1)	Variant	91.8%

# Classification Online Learning



# Dynamic Classifiers

- Automatically classify after new samples have been received
  - A maximum of 2000 (default) samples can be stored
- You can turn this option off if you wish to control which samples are used
  - Samples can be downloaded and deleted from the server
  - Downloaded samples can be added to the classification set

